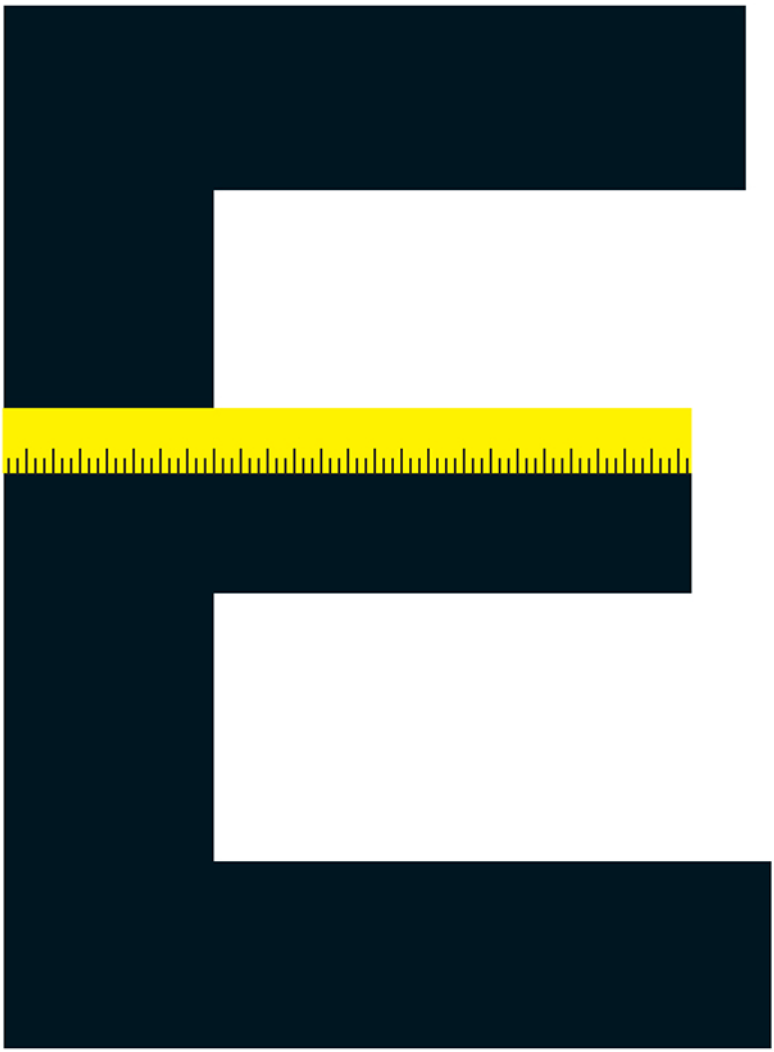
El camino entre el efecto y la causa

# DOMINAR LA ECONOMETRÍA



Joshua D. Angrist
Jörn-Steffen Pischke

# DOMINAR LA ECONOMETRÍA

# DOMINAR LA ECONOMETRÍA

## El camino que va de la causa al efecto

Joshua D. Angrist
y Jörn-Steffen Pischke

Traducción de
Dulcinea Otero-Piñeiro

Revisión científico-técnica de
David Galadí-Enríquez

Antoni Bosch editor

Antoni Bosch editor, S.A.
Palafolls 28, 08017 Barcelona, España
Tel. (+34) 93 206 07 30
info@antonibosch.com
www.antonibosch.com

Título original de la obra: *Mastering Metrics*

Copyright © 2015 Princeton University Press
© 2016 de la edición en español: Antoni Bosch editor, S.A.

ISBN: 978-84-941595-0-3
Depósito legal: B. 10.806-2016

Diseño de la cubierta: Compañía
Maquetación: JesMart
Corrección: Andreu Navarro
Impresión: Bookprint

Impreso en España
*Printed in Spain*

Cualquier forma de reproducción, distribución, comunicación pública o transformación de esta obra solo puede ser realizada con la autorización de sus titulares, salvo excepción prevista por la ley. Diríjase a CEDRO (Centro Español de Derechos Reprográficos, www.cedro.org) si necesita fotocopiar o escanear algún fragmento de esta obra.

# Índice

Introducción ..... 9

1 Experimentos aleatorios ..... 15
1.1 En la salud y en la enfermedad ..... 15
1.2 El rastro de Oregón ..... 39
Maestros de la econometría: de Daniel a R. A. Fisher ..... 46
Apéndice: Dominar la inferencia ..... 48

2 Regresión ..... 63
2.1 Historia de dos universidades ..... 64
2.2 Emparéjame y hazme una regresión ..... 73
2.3 *¿Ceteris paribus?* ..... 86
Maestros de la econometría: Galton y Yule ..... 98
Apéndice: Teoría de la regresión ..... 100

3 Variables instrumentales ..... 119
3.1 El dilema *charter* ..... 120
3.2 Contra el abuso ..... 137
3.3 La bomba poblacional ..... 146
Maestros de la econometría: El increíble equipo de los Wright ..... 163
Apéndice: Teoría de variables instrumentales ..... 166

4 Diseños de regresión discontinua ..... 171
4.1 Cumpleaños y funerales ..... 172
4.2 La ilusión de la élite ..... 189
Maestros de la econometría: Donald Campbell ..... 201

5 Diferencias en diferencias ..... 205
5.1 El experimento Misisipi ..... 206
5.2 Bebe y vive ..... 219
Maestros de la econometría: John Snow ..... 232
Apéndice: Errores típicos en regresiones DD ..... 233

6 El valor de la enseñanza ..... 237
6.1 Formación académica, experiencia e ingresos ..... 237
6.2 Los gemelos doblan la diversión ..... 246
6.3 Econometristas: por sus instrumentos los conocerás ..... 252
6.4 Flamantes badanas en el estado de la estrella solitaria ..... 265
Apéndice: Sesgo debido a errores de medida ..... 269

Abreviaturas y acrónimos ..... 275

Notas empíricas ..... 279

Relación de figuras ..... 299

Relación de tablas ..... 301

Agradecimientos ..... 303

Índice analítico ..... 305

# Introducción

Maestro ciego Po: Cierra los ojos. ¿Qué oyes?
Joven Kwai Chang Caine: Oigo el agua, oigo los pájaros.
Maestro Po: ¿No oyes el latido de tu corazón?
Pequeño saltamontes Kwai Chang Caine: No.
Maestro Po: ¿Oyes el saltamontes que hay a tus pies?
Pequeño saltamontes Kwai Chang Caine: Anciano, ¿cómo es posible que oigas esas cosas?
Maestro Po: Joven, ¿cómo es posible que tú no?

Episodio piloto de *Kung Fu*

Los economistas no se merecen su mala fama. La economía es tan apasionante como pueda serlo cualquier otra ciencia: el mundo es nuestro laboratorio, y la inmensa diversidad de personas que lo habitan conforma nuestro objeto de estudio.

Lo interesante de nuestro trabajo proviene de la oportunidad que brinda para aprender sobre causas y efectos en el quehacer humano. Las grandes cuestiones del día a día son *nuestras* cuestiones: ¿Deparará crecimiento económico cierta política monetaria, o servirá tan sólo para atizar el fuego de la inflación? Es lo que desearían saber los granjeros de Iowa y quien ostenta la presidencia de la Reserva Federal. ¿De verdad conseguirá el seguro médico obligatorio que los estadounidenses estén más sanos? Esta yesca política incendia las tertulias radiofónicas. Nosotros abordamos estas cuestiones con frialdad, armados con argumentos no apasionados, sino numéricos.

En el campo de la econometría aplicada los economistas emplean datos para resolver problemas de causa y efecto. Las herramientas del oficio econométrico son el análisis disciplinado de datos combinado con la maquinaria de la inferencia estadística. Pero nuestra labor también cuenta con una vertiente mística: buscamos la verdad, pero la verdad no se revela íntegra, y los mensajes que transmiten los datos exigen interpretación. Para ello buscamos inspiración en el viaje de Kwai Chang Caine, protagonista de la mítica serie de televisión *Kung Fu*. Caine, un monje shaolin mestizo, recorre el Oeste norteamericano del siglo XIX en busca de su medio hermano nacido en Estados Unidos. Durante la búsqueda, Caine cuestiona todo lo que ve en los asuntos humanos, lo cual lo lleva a descubrir relaciones ocultas y explicaciones profundas. Igual que el viaje de Caine, la senda de la econometría está iluminada por las preguntas.

## **Ceteris paribus** *o lo demás permanece constante*

De acuerdo con una alarmante tendencia de la que tal vez haya oído hablar, la proporción de estudiantes estadounidenses que completan sus estudios universitarios en el plazo de tiempo estipulado ha dado un giro brusco a peor. Los políticos y analistas políticos culpan del descenso en los índices de graduaciones universitarias a una combinación nefasta de subida de tasas y los grandes préstamos que afrontan muchos alumnos para financiar sus estudios. Puede que el aumento de los préstamos estudiantiles descarríe a algunos que en caso contrario seguirían por buen camino. El hecho de que los estudiantes con más probabilidad de abandono escolar sean los que suelen soportar préstamos de mayor cuantía parecería confirmar esta hipótesis.

Si pudiéramos, preferiríamos pagar los estudios con riquezas heredadas en lugar de hacerlo con dinero prestado. Sin embargo, tal como analizaremos en detalle, es probable que la formación potencie lo suficiente los ingresos como para que la mayoría de los graduados pueda devolver el préstamo. Entonces ¿cómo debería interpretarse la correlación negativa que existe entre el importe de la deuda y los índices de graduaciones universitarias? ¿Es que el endeudamiento *provoca* la deserción de los deudores? El primer interrogante que de-

bemos plantearnos en este contexto es quién solicita préstamos de mayor cuantía. Los estudiantes que piden préstamos más altos suelen provenir de familias con ingresos medios o bajos, puesto que las familias más ricas disponen de más ahorros. Por muchas razones, los estudiantes procedentes de familias con menos ingresos tienen menos probabilidades de concluir los estudios universitarios que los alumnos procedentes de familias con ingresos más altos, con independencia de si se endeudan mucho o poco. Por tanto, deberíamos tomarnos con escepticismo la afirmación de que las deudas más altas son la causa de la menor proporción de finalización a tiempo de los estudios, cuando esa afirmación se basa únicamente en comparar los índices de finalización de estudios entre quienes tienen deudas más o menos elevadas. Debido a la correlación entre situación familiar de partida y deuda estudiantil, el contraste en los índices de graduación entre quienes tienen o no préstamos ya no es una comparación donde *todo lo demás permanece constante*.

Como universitarios especializados en economía, aprendimos por primera vez esa idea de *todo lo demás permanece constante* a partir de su nombre en latín, *ceteris paribus*. Las comparaciones que se establecen en condiciones *ceteris paribus* admiten una interpretación causal. Imaginemos dos estudiantes idénticos en todo, de tal manera que sus familias cuenten con los mismos recursos económicos y sus padres tengan un nivel de formación similar. Uno de estos estudiantes prácticamente gemelos se paga los estudios universitarios mediante un préstamo y el otro con ahorros. Como son iguales en todo lo demás (ambos recibieron de sus abuelas unos pequeños ahorros), las diferencias en cuanto a logros académicos sólo son atribuibles al hecho de que uno de ellos pidió un préstamo. Hasta el día de hoy nos preguntamos por qué hay tantos estudiantes de economía que se topan por primera vez con esta idea crucial en forma de expresión latina; quizá sea una conspiración para evitar que reflexionen sobre ella. Porque, tal como deja entrever esta comparación hipotética, las comparaciones reales sujetas al supuesto de que *todo lo demás permanece constante* son difíciles de construir, y habrá quien lo considere incluso *impossibile* (que es italiano, no latín, pero al menos aún se habla).

Difíciles de construir, tal vez, pero no necesariamente imposibles. El arte de la econometría usa datos para conseguir condiciones en las que *el resto permanece constante* a pesar de los obstáculos (lo que

se llama *sesgo de selección* o *sesgo por variables omitidas*) que surgen a lo largo del recorrido que lleva desde los números brutos hasta el conocimiento causal fiable. El camino hacia el discernimiento causal es accidentado y se ensombrece a medida que serpentea entre los escollos del sesgo de selección. Sin embargo, los maestros de la econometría caminan por esta senda con confianza y, al mismo tiempo, con humildad para enlazar con éxito causa y efecto.

La primera línea de actuación para abordar el problema de la causalidad consiste en el experimento aleatorio. En un experimento aleatorio, los investigadores cambian las variables causales de interés (por ejemplo, la disponibilidad de ayuda financiera universitaria) de un grupo seleccionado usando algo parecido al lanzamiento de una moneda al aire. Al cambiar las circunstancias de manera aleatoria, se vuelve altamente probable que la variable de interés no guarde relación con los numerosos factores adicionales que condicionan los resultados que queremos estudiar. La asignación aleatoria no es lo mismo que mantener fijo todo lo demás, pero ejerce el mismo efecto. La manipulación aleatoria hace que el principio de *el resto permanece constante* se cumpla, en promedio, tanto en el grupo que fue alterado como en el que no. Tal como se explica en el capítulo 1, «en promedio» suele ser suficiente.

Los experimentos aleatorios son una de nuestras herramientas prioritarias. Por desgracia, los experimentos sociales aleatorios son caros y pueden tardar en dar frutos, mientras que los fondos para investigar son escasos y la vida es corta. De ahí que a menudo los maestros de la econometría recurran a proyectos de investigación menos potentes pero más accesibles. Sin embargo, cuando los experimentos aleatorios no sean practicables, seguiremos soñando con los experimentos que nos gustaría hacer. El concepto de experimento ideal ayuda a mantener el rigor en la investigación econométrica. Este volumen pone de manifiesto de qué modo la aplicación inteligente de nuestras cinco herramientas econométricas preferidas potencia al máximo la capacidad de un experimento real para revelar la causalidad.

Ilustramos nuestras cinco armas econométricas preferidas mediante una serie de estudios econométricos bien elaborados y relevantes. Estos estudios de efectos causales gozan del visto bueno del Gran Maestro Oogway del Palacio de Jade de *Kung Fu Panda* y se caracterizan por su magnificencia. Los métodos empleados (asignación aleatoria, regresión, variables instrumentales, diseños de regresión discontinua, y diferencias en diferencias) integran el grupo de los Cinco Furiosos de la investigación econométrica. Para quienes se inicien en la materia, y debido al debate que existe en la actualidad en Estados Unidos en relación con la asistencia médica, el primer capítulo describe dos experimentos sociales que evidencian si, tal como creen muchos políticos, los seguros médicos contribuyen a mejorar la salud de quienes los contratan. Del capítulo 2 al 5 pondremos en funcionamiento el resto de herramientas para dar respuesta a cuestiones relevantes que van desde las ventajas de estudiar en universidades privadas y en escuelas *charter*, hasta el coste del consumo de alcohol entre adolescentes, y los efectos de las inyecciones de liquidez por parte de los bancos centrales.

El último capítulo pone a prueba a los Cinco Furiosos regresando al ámbito de la educación. En promedio, en Estados Unidos, las personas con estudios universitarios ganan alrededor del doble que las que cuentan con estudios medios, una brecha salarial que parece ir en aumento. En el capítulo 6 nos planteamos si esa diferencia es indicativa de una gran rentabilidad causal de la formación académica o tan sólo es un reflejo de las numerosas ventajas adicionales con que cuentan quienes tienen más estudios (como unos padres más instrui-

dos). ¿Se puede valorar la relación entre la formación académica y los ingresos en condiciones *ceteris paribus*, o siempre nos bloquearán el camino los escollos del sesgo de selección? El reto de cuantificar la relación causal entre la formación y los ingresos supone una prueba de fuego apasionante para las armas de la econometría y los maestros que las empuñan.

1

# Experimentos aleatorios

Kwai Chang Caine: Lo que sucede en la vida de un hombre ya está escrito. El hombre debe andar por la vida como disponga su destino.

Anciano: Pero cada cual es libre de vivir como elija. Aunque parezcan opuestas, ambas afirmaciones son verdad.

### *Nuestro camino*

Nuestro camino comienza con la *asignación aleatoria*, como base para abordar cuestiones causales y como referente para valorar los resultados obtenidos por otros métodos. Ilustraremos el poder de una asignación aleatoria mediante dos análisis aleatorios de los efectos de los seguros médicos. El apéndice de este capítulo también emplea una base experimental para presentar los conceptos y métodos de la inferencia estadística.

## 1.1 En la salud y en la enfermedad

La Ley de Atención Médica Asequible (*Affordable Care Act*, ACA) ha demostrado ser una de las novedades políticas más controvertidas e interesantes que hemos visto en Estados Unidos. La ACA obliga a los estadounidenses a contratar un seguro médico, y penaliza a través de los impuestos a quien no lo contrata de forma voluntaria. La cues-

tión de cuál deba ser la verdadera función del gobierno dentro del mercado de la asistencia sanitaria tiene muchas vertientes. Una de ellas es el efecto causal de los seguros médicos en la salud. Estados Unidos gasta una proporción mayor de su producto interior bruto en atención sanitaria que otros países desarrollados y, a pesar de ello, sus habitantes no gozan de buena salud. Por ejemplo, los estadounidenses tienen más probabilidad de padecer sobrepeso, y de morir antes, que sus primos canadienses, quienes dedican a este fin tan sólo unas dos terceras partes de lo que se gasta en Estados Unidos, y este país también se diferencia de otros países desarrollados en que no posee un seguro médico universal. Tal vez detrás de todo esto haya una relación causal.

Los mayores estadounidenses están cubiertos por un programa federal llamado Medicare, mientras que parte de la población sin recursos (que incluye a la mayoría de las madres solteras, sus hijos y muchos otros niños sin recursos) está cubierta por Medicaid. Muchos de los trabajadores pobres más jóvenes han estado mucho tiempo sin asegurar. De hecho, muchos estadounidenses sin seguro han optado por no participar en el plan de seguros que proporciona su empleador.[1] Estos trabajadores cuentan, tal vez con acierto, con los servicios de urgencias de los hospitales, que no se pueden negar a atenderlos, para resolver sus necesidades de atención sanitaria. Pero los servicios de urgencias tal vez no sean los mejores sitios para tratar, por ejemplo, una gripe, o para controlar afecciones crónicas como la diabetes y la hipertensión, ambas muy generalizadas entre los estadounidenses sin recursos. La unidad de urgencias no está obligada a ofrecer cuidados a largo plazo. Por tanto, es evidente que el seguro médico impuesto por el gobierno podría reportar mejoras en la salud. El empeño por implantar un seguro médico universal se debe en parte al convencimiento de que así será.

El principio *ceteris paribus* en este contexto compara la salud de alguien cubierto por un seguro, con la salud de esa misma persona si no dispusiera de asistencia médica gratuita (aparte de la que brindan los servicios de urgencias). Este cotejo pone de relieve un problema empírico fundamental: o la gente está asegurada, o no lo está. No

[1] Para conocer más a fondo este hecho sorprendente, véase Jonathan Gruber, «Covering the Uninsured in the United States», *Journal of Economic Literature*, vol. 46, n.º 3, septiembre de 2008, páginas 571-606.

puede estar en ambas situaciones, al menos no al mismo tiempo y exactamente en las mismas circunstancias.

En su conocido poema titulado «The Road Not Taken» [«El camino que no tomé»], Robert Frost usó la metáfora de un cruce de caminos para describir los efectos causales de cada decisión personal:

> En un bosque amarillo divergían dos senderos
> pero era imposible elegir los dos
> por ser yo sólo uno, y de pie con esmero
> contemplé la apariencia que mostraba el primero,
> hasta donde torcía en la vegetación.

El viajero de Frost concluye diciendo:

> En un bosque amarillo divergían dos caminos
> de los cuales tomé el menos concurrido
> y esa opción me marcó ya todo lo demás.

El viajero afirma que aquella elección fue decisiva pero, al ser una sola persona decidiendo, no puede estar seguro de ello. Un viaje posterior o el testimonio de otros viajeros tampoco le aclararía nada. Puede que nuestro narrador fuera mayor y más sabio la segunda vez, y que otros viajeros tuvieran vivencias distintas tomando ese camino. Lo mismo sucede con cualquier elección, incluidas las relacionadas con los seguros médicos: ¿estaría sano un hombre con una afección cardiaca y sin seguro si tuviera seguro? En la novela *Años luz*,[2] de James Salter, el indeciso narrador observa: «Los actos destruyen sus alternativas, esa es la paradoja». No podemos saber qué hay al final de un camino que no se tomó.

En efecto, no se puede, pero hay ciertos indicios que permiten estudiar el asunto. Este capítulo presenta algunas pistas sobre caminos relacionados con seguros médicos. El punto de partida es la Encuesta Nacional de Salud (National Health Interview Survey, NHIS), que se realiza cada año entre la población estadounidense y proporcio-

[2] James Salter, *Años luz*, Salamandra, Barcelona, 2013, trad. de Jesús Zulaika. *(N. de la T.)*

na una información detallada sobre salud y seguros médicos. Entre muchas otras cosas, esta encuesta pregunta: «En términos generales, ¿diría usted que tiene una salud excelente, muy buena, buena, regular o mala?». Usamos esta pregunta para codificar un indicador que asigna un 5 a la salud excelente y un 1 a la mala salud de una muestra de gente casada –tanto con seguro médico como sin él–[3] que respondió la NHIS en 2009. Ese índice es nuestro *resultado*: una medida que nos interesa estudiar. La relación causal que ahora nos interesa está determinada por una variable que indica la cobertura mediante seguros médicos privados. A esa variable la hemos denominado el *tratamiento*, término que tomamos prestado de los textos especializados de ensayos clínicos, aunque los tratamientos que nos interesan a nosotros no tienen por qué ser médicos, como son los fármacos o las intervenciones quirúrgicas. En este contexto, las personas aseguradas pueden considerarse el *grupo experimental* o *de tratamiento*, y las que no disponen de seguro conforman el *grupo de control* o *de comparación*. Un buen grupo de control revela el destino que habrían tenido las personas tratadas si vivieran en un mundo contrafactual en el que no hubieran sido tratadas.

La primera fila de la tabla 1.1 compara la media de la puntuación de salud de estadounidenses asegurados y no asegurados, donde las estadísticas se tabulan por separado para maridos y esposas.[4] En efecto, quienes tienen un seguro médico están más sanos que los que no lo tienen, una brecha aproximada de 0,3 puntos en el baremo masculino y de 0,4 puntos en el femenino. Estas diferencias son grandes cuando se comparan con la desviación típica de las puntuaciones, que viene a ser de 1. (Las desviaciones típicas, que se dan entre corchetes en la tabla 1.1, miden la variabilidad de los datos. En el apéndice de este capítulo se analiza la fórmula correspondiente.) Estas grandes brechas revelan la mejora de la salud que estamos buscando.

[3] Nuestra muestra incluye individuos de edades comprendidas entre dieciséis y cincuenta y nueve años que, por tanto, aún no pertenecen a Medicare.

[4] En el apartado titulado «Notas empíricas», que figura después del último capítulo, se da información detallada sobre esta tabla y la mayor parte del resto de tablas y figuras del libro.

### *Comparaciones fructuosas e infructuosas*

Las comparaciones simples, como las que aparecen en la parte superior de la tabla 1.1, suelen citarse como indicio de efectos causales. Sin embargo, no pocas veces estas comparaciones resultan engañosas. Una vez más el problema radica en si de verdad *el resto permanece constante* o no. Comparar a gente que tiene y no tiene seguro médico no es comparar manzanas con manzanas, sino manzanas con naranjas, o algo peor.

Entre otras diferencias, quienes tienen seguro médico cuentan con una formación académica más elevada, mayores ingresos, y más probabilidades de trabajar, que quienes carecen de él. Esto se ve en el apartado B de la tabla 1.1, el cual muestra las características promedio de los encuestados en la NHIS que tienen o no seguro médico. Muchas de las diferencias que figuran en la tabla son grandes (por ejemplo, una brecha de casi tres años en escolaridad); la mayoría son lo bastante precisas desde un punto de vista estadístico como para descartar la hipótesis de que esas discrepancias se deban al azar (consúltese el apéndice del capítulo para repasar la significancia estadística). No le sorprenderá saber que la mayoría de las variables tabuladas tienen una correlación elevada con la salud, así como con la tenencia o no de seguro médico. Por ejemplo, la gente con más formación académica suele estar más sana y tener más representación dentro del grupo de los asegurados. Esto podría deberse a que la gente más formada practica más ejercicio, fuma menos y tiende a usar más el cinturón de seguridad. Es evidente que la diferencia de salud entre los encuestados en la NHIS asegurados y no asegurados refleja, al menos en parte, la ventaja en cuanto a formación de los asegurados.

Nuestro esfuerzo por descubrir una relación causal entre seguros y salud se apoyó en el desarrollo de la metáfora de los dos caminos que emplea Frost. Usamos la letra $Y$ para representar la salud, el resultado o variable de interés. Para marcar con claridad cuándo hablamos de una persona concreta, usamos subíndices en sustitución de sus nombres: $Y_i$ es la salud del individuo $i$. El resultado $Y_i$ aparece registrado en nuestros datos. A la hora de elegir si paga o no un seguro médico, la persona $i$ tiene dos *resultados potenciales*, pero sólo se tiene en cuenta uno de ellos. Para diferenciar un resultado potencial de otro añadimos un segundo subíndice: si se elige el camino sin seguro médico, tenemos $Y_{0i}$ (léase «y-cero-i») para la persona $i$, mientras que

el camino con seguro médico da lugar a $Y_{1i}$ (léase «y-uno-i») para la persona *i*. Los resultados potenciales se encuentran al final del camino que tome cada cual. El efecto causal del seguro en la salud es la diferencia entre ambas posibilidades, que se escribe como $Y_{1i} - Y_{0i}$.[5]

**Tabla 1.1. Características demográficas y estado de salud de parejas aseguradas y no aseguradas en la NHIS[6]**

| | Maridos | | | Esposas | | |
|---|---|---|---|---|---|---|
| | Algún SM | Ningún SM | Diferencia | Algún SM | Ningún SM | Diferencia |
| **A. Estado de salud** | | | | | | |
| Puntuación de salud | 4.01 | 3.70 | .31 | 4.02 | 3.62 | .39 |
| | [.93] | [1.01] | (.03) | [.92] | [1.01] | (.04) |
| **B. Características** | | | | | | |
| No blancos | .16 | .17 | –.01 | .15 | .17 | –.02 |
| | | | (.01) | | | (.01) |
| Edad | 43.98 | 41.26 | 2.71 | 42.24 | .39.62 | 2.62 |
| | | | (.29) | | | (1.30) |
| Formación académica | 14.31 | 11.56 | 2.74 | 14.44 | 11.80 | 2.64 |
| | | | (.10) | | | (.11) |
| Tamaño familiar | 3.50 | 3.98 | –.47 | 3.49 | 3.93 | –.43 |
| | | | (.05) | | | (.05) |
| Con trabajo | .92 | .85 | .07 | .77 | .56 | .21 |
| | | | (.01) | | | (.02) |
| Ingresos familiares | 106,467 | 45,656 | 60,810 | 106,212 | 46,385 | 59,828 |
| | | | (1,355) | | | (1,406) |
| Tamaño de la muestra | 8,114 | 1,281 | | 8264 | 1,131 | |

*Notas*: Esta tabla contiene datos sobre características promedio de parejas casadas aseguradas y no aseguradas que respondieron la Encuesta Nacional de Salud (NHIS) de 2009. Las columnas (1), (2), (4) y (5) muestran características promedio del grupo de individuos especificado en el encabezamiento de cada columna. Las columnas (3) y (6) dan la diferencia entre la característica promedio de individuos con y sin seguro médico (SM). Las desviaciones típicas figuran entre corchetes; los errores típicos se dan entre paréntesis.

[5] A pesar de las consideraciones de Robert Frost, la econometría no es poesía. Un mínimo de notación matemática nos permite describir y tratar con precisión relaciones sutiles. Asimismo recurrimos a la cursiva para introducir términos que se usan con frecuencia, como *resultados potenciales*, que tienen un significado especial para los maestros de la econometría.

[6] En el texto de esta obra, el lector encontrará los números decimales escritos con coma y las cifras superiores al millar, con punto. Sin embargo, en las tablas se mantienen los datos tal cual los muestran los programas informáticos de uso habitual en econometría; es decir, se reserva el punto para indicar decimales y la coma para los millares. *(N. del Ed.)*

Concretando un poco más, consideremos la historia del estudiante Khuzdar Khalat, llegado recientemente de Kazajistán como alumno visitante al Massachusetts (Institute of Technology MIT). Kazajistán cuenta con un sistema nacional de asistencia sanitaria que cubre de forma automática a toda la población (aunque no iríamos allí únicamente por su asistencia sanitaria). Al llegar a Cambridge, Khuzdar se entera con sorpresa de que los alumnos del MIT deben decidir si pertenecer o no al plan de seguro médico que ofrece la universidad y para el que el MIT impone una tasa considerable. Tras meditarlo, Khuzdar concluye que vale la pena pagar el seguro del MIT porque teme que los frescos aires de Nueva Inglaterra le causen alguna infección en las vías respiratorias. Digamos que $Y_{0i} = 3$ y que $Y_{1i} = 4$, donde $i$ = Khuzdar. Para él, el efecto causal de la contratación del seguro supone subir un puesto en la escala del NHIS:

$$Y_{1,\text{Khuzdar}} - Y_{0,\text{Khuzdar}} = 1.$$

La tabla 1.2 resume esta información.

**Tabla 1.2. Resultados y tratamientos para Khuzdar y María**

| | Khuzdar Khalar | María Moreno |
|---|---|---|
| Resultado potencial sin seguro $Y_{0i}$ | 3 | 5 |
| Resultado potencial con seguro $Y_{1i}$ | 4 | 5 |
| Tratamiento (opción elegida en cuanto a seguro) $D_i$ | 1 | 0 |
| Resultado real en cuanto a estado de salud $Y_i$ | 4 | 5 |
| Efecto del tratamiento $Y_{1i} - Y_{0i}$ | 1 | 0 |

Conviene hace hincapié en que la tabla 1.2 es imaginaria: algunas de las informaciones que describe son imposibles de conocer. Khuzdar puede contratar el seguro, lo que revela el valor $Y_{1i}$, o no hacerlo, en cuyo caso se revela el valor $Y_{0i}$. Khuzdar ha transitado por muchos caminos largos y polvorientos de Kazajistán, pero aun así no puede estar seguro de qué hay al final de las sendas que no tomó.

María Moreno también asistirá al MIT este año; ella es del altiplano andino de Chile. Los inviernos de Boston preocupan poco a María porque es una persona fuerte que no enferma con facilidad, así que renuncia al seguro del MIT con la intención de dedicar el dinero a

viajar. Como María tiene $Y_{0,\text{María}} = Y_{1,\text{María}} = 5$, el efecto causal del seguro en su salud es

$$Y_{1,\text{María}} - Y_{0,\text{María}} = 0.$$

Los números de María también figuran en la tabla 1.2.

Como Khuzdar y María eligen opciones distintas en relación con el seguro médico, ofrecen una comparación interesante. La salud real de Khuzdar es $Y_{\text{Khuzdar}} = Y_{1,\text{Khuzdar}} = 4$, mientras que la de María es $Y_{\text{María}} = Y_{0,\text{María}} = 5$. La diferencia entre ellos es

$$Y_{\text{Khuzdar}} - Y_{\text{María}} = -1.$$

Tomada al pie de la letra, esta cantidad (que observamos nosotros) sugiere que la decisión de Khuzdar de contratar el seguro es contraproducente. A pesar de la cobertura médica que le ofrece su seguro del MIT, la salud asegurada de Khuzdar es peor que la no asegurada de María.

De hecho, la comparación entre el frágil Khuzdar y la robusta María revela bien poco sobre los efectos causales de sus decisiones. Esto se ve al relacionar los resultados observados y los potenciales del siguiente modo:

$$\begin{aligned} Y_{\text{Khuzdar}} - Y_{\text{María}} &= Y_{1,\text{Khuzdar}} - Y_{0,\text{María}} \\ &= \underbrace{Y_{1,\text{Khuzdar}} - Y_{0,\text{Khuzdar}}} + \underbrace{\{Y_{0,\text{Khuzdar}} - Y_{0,\text{María}}\}} \end{aligned}$$

La segunda línea de esta ecuación deriva de sumar y restar $Y_{0,\text{Khuzdar}}$, con lo que se obtienen dos comparaciones ocultas que determinan la que vemos. La primera comparación, $Y_{1,\text{Khuzdar}} - Y_{0,\text{Khuzdar}}$, es el efecto causal del seguro médico en Khuzdar, y vale 1. La segunda, $Y_{0,\text{Khuzdar}} - Y_{0,\text{María}}$, es la diferencia en cuanto a estado de salud entre ambos estudiantes si los dos decidieran no contratar el seguro. Este término, que vale –2, refleja la fragilidad de Khuzdar. Dentro del contexto de nuestro empeño por destapar efectos causales, la falta de equivalencia reflejada por la segunda expresión se denomina *sesgo de selección*.

Tal vez crea usted que el sesgo de selección tiene algo que ver con el hecho de centrarnos en individuos particulares en lugar de trabajar con grupos, donde quizá cabría esperar que las diferencias no relevantes «se compensaran» en promedio. Pero el difícil problema del sesgo de selección se mantiene en las comparaciones de grupos, aunque, en lugar de efectos causales individuales, centremos la aten-

ción en *efectos causales medios*. En un grupo de $n$ personas, los efectos causales medios se escriben $Avg_n[Y_{1i} - Y_{0i}]$, donde el promedio (o media) se obtiene de la manera habitual (es decir, se suman los resultados individuales y se dividen entre $n$):

$$Avg_n[Y_{1i} - Y_{0i}] = \frac{1}{n}\sum_{i=1}^{n}[Y_{1i} - Y_{0i}]$$

$$= \frac{1}{n}\sum_{i=1}^{n} Y_{1i} - \frac{1}{n}\sum Y_{0i}. \qquad (1.1)$$

El símbolo $\sum_{i=1}^{n}$ indica una suma de todo el mundo desde $i = 1$ hasta $n$, donde $n$ es el tamaño del grupo que estamos promediando. Nótese que los dos sumatorios de la ecuación (1.1) se aplican a todos los individuos del grupo de interés. El efecto causal medio del seguro médico compara la salud media en situaciones hipotéticas donde todas las personas del grupo tienen y no tienen seguro médico. Desde un punto de vista computacional, esta es la media de efectos causales individuales como $Y_{1,\text{Khuzdar}} - Y_{0,\text{Khuzdar}}$ e $Y_{1,\text{María}} - Y_{0,\text{María}}$ para cada estudiante de nuestros datos.

Un análisis del efecto causal medio del seguro médico empezará, como es natural, comparando la salud media de grupos de personas aseguradas y no aseguradas, como en la tabla 1.1. Este cotejo lo facilita la construcción de una *variable binaria*, $D_i$, que toma los valores 0 y 1 para indicar si se está asegurado o no:

$$D_i = \begin{cases} 1 & \text{si } i \text{ tiene seguro} \\ 0 & \text{en caso contrario.} \end{cases}$$

Ahora podemos escribir $Avg_n[Y_i|D_i = 1]$ para la media entre los asegurados, y $Avg_n[Y_i \,|\, D_i = 0]$ para la media entre los no asegurados. Esas cantidades son medias *condicionadas* de la situación en cuanto a seguro médico.[7]

[7] Ordene las $n$ observaciones de $Y_i$ de forma que las $n_0$ observaciones del grupo indicadas por $D_i = 0$ precedan las $n_1$ observaciones del grupo $D_i = 1$. La media condicionada

$$Avg_n[Y_i \,|\, D_i = 0] = \frac{1}{n_0}\sum_{i=1}^{n_0} Y_i$$

es la media muestral para las $n_0$ observaciones en el grupo $D_i = 0$. La expresión $Avg_n[Y_i|D_i = 1]$ se calcula de forma análoga a partir de las $n_1$ observaciones restantes.

La media $Y_i$ de los asegurados será necesariamente una media del resultado $Y_{1i}$, pero no contiene ninguna información sobre $Y_{0i}$. Del mismo modo, la media $Y_i$ de los no asegurados es una media del resultado $Y_{0i}$, pero esta media carece de información sobre la correspondiente $Y_{1i}$. En otras palabras, el camino tomado por quienes están asegurados acaba llegando a $Y_{1i}$, mientras que la senda seguida por quienes no tienen seguro lleva hasta $Y_{0i}$. Esto conduce a su vez a una conclusión sencilla pero importante sobre la diferencia en cuanto a salud media debida a si se tiene o no seguro médico:

$$\begin{aligned} &\textit{Diferencia de medias de grupo} \\ &= Avg_n[Y_i \mid D_i = 1] - Avg_n[Y_i \mid D_i = 0] \\ &= Avg_n[Y_{1i} \mid D_i = 1] - Avg_n[Y_{0i} \mid D_i = 0], \end{aligned} \tag{1.2}$$

una expresión que recalca el hecho de que las comparaciones de la tabla 1.1 revelan algo sobre resultados potenciales, pero no necesariamente lo que queremos saber. Nosotros buscamos $Avg_n[Y_{1i} - Y_{0i}]$, un efecto causal medio relacionado con el resultado $Y_{1i}$ de todos y el resultado $Y_{0i}$ de todos, pero vemos el resultado medio $Y_{1i}$ tan sólo para los asegurados, y el $Y_{0i}$ medio solamente de los no asegurados.

Para entender mejor la ecuación (1.2) servirá de ayuda imaginar que el seguro médico mejora la salud de todos en una cantidad constante, $\kappa$. Como acostumbramos a hacer en el gremio, usamos letras griegas para etiquetar esos *parámetros* y, de este modo, distinguirlos de las variables o los datos; en este caso usamos la letra kappa. La *hipótesis de efectos constantes* nos permite escribir:

$$Y_{1i} = Y_{0i} + \kappa, \tag{1.3}$$

o bien $Y_{1i} - Y_{0i} = \kappa$, que es equivalente. En otras palabras, $\kappa$ es el efecto causal tanto individual como medio del seguro médico sobre la salud. La cuestión es cómo se relacionan con $\kappa$ comparaciones como las que aparecen en la parte superior de la tabla 1.1.

Empleando el modelo de efectos constantes (ecuación (1.3)) para sustituir $Avg_n[Y_i|D_i = 1]$ en la ecuación (1.2), tenemos:

$$
\begin{aligned}
&Avg_n[Y_i \mid D_i = 1] - Avg_n[Y_{0i} \mid D_i = 0] \\
&\quad = \{\kappa + Avg_n[Y_{0i} \mid D_i = 1] - Avg_n[Y_{0i} \mid D_i = 0] \\
&\quad = \kappa + \{Avg_n[Y_{0i} \mid D_i = 1] - Avg_n[Y_{0i} \mid D_i = 0]. \qquad (1.4)
\end{aligned}
$$

Esta ecuación revela que las comparaciones de salud entre quienes tienen seguro y quienes no lo tienen equivalen al efecto causal de interés *(κ)*, más la diferencia entre los valores medios de $Y_{0i}$ para asegurados y no asegurados. Igual que en la parábola de Khuzdar y María, esta segunda expresión describe un sesgo de selección. En concreto, la diferencia de salud media debida a la situación en cuanto al seguro se puede escribir así:

*Diferencia de medias de grupo*
*= efecto causal medio + sesgo de selección*,

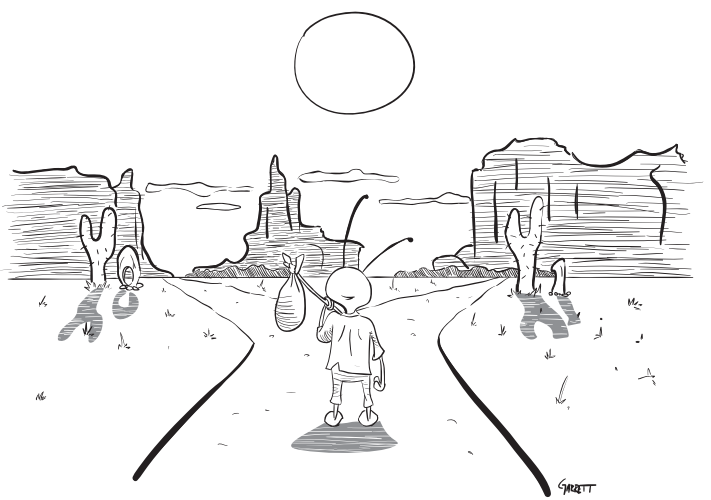
donde el sesgo de selección se define como la diferencia en $Y_{0i}$ media entre los grupos que se están comparando.

¿Cómo sabemos que la diferencia en el promedio debida a la situación en cuanto a seguro médico está contaminada por el sesgo de selección? Lo sabemos porque $Y_{0i}$ resume todo lo relacionado con la salud sobre la persona *i*, salvo si está asegurada o no. La parte inferior de la tabla 1.1 documenta diferencias importantes no relacionadas con el seguro médico entre asegurados y no asegurados, lo que revela que aquí *ceteris* no es *paribus* en numerosos aspectos. Los asegurados de la Encuesta Nacional de Salud están más sanos por todo tipo de razones, que tal vez incluyan los efectos causales del seguro médico. Pero los asegurados también están más sanos porque tienen más formación, entre otras cosas. Para ver por qué es esto relevante, imagine un mundo en el que el efecto causal de estar asegurado fuera cero (es decir, $\kappa = 0$). Hasta en un mundo así cabría esperar que los encuestados en la NHIS provistos de seguro médico estuvieran más sanos, por la sencilla razón de que tienen más formación, más ingresos económicos, etcétera.

Cerramos este debate haciendo hincapié en la sutil relevancia de la información contenida en el apartado B de la tabla 1.1. En él se aprecia que los grupos comparados difieren en cuanto a aspectos observables. Tal como veremos en el próximo capítulo, si la única

fuente del sesgo de selección fuera un conjunto de diferencias en cuanto a características que se pueden observar y medir, el sesgo de selección resulta (bastante) sencillo de eliminar. Supongamos, por ejemplo, que el único origen del sesgo de selección al comparar el asunto del seguro médico fuera el nivel de estudios. Este sesgo se eliminaría en cuanto se tomaran muestras de gente con la misma formación académica como, por ejemplo, graduados universitarios. En esa muestra tanto los asegurados como los no asegurados tendrán el mismo nivel de formación, porque éste es el mismo para todas las personas de la muestra.

La sutileza de la tabla 1.1 surge porque en las situaciones en las que proliferan las diferencias observables, también deberían aumentar nuestras sospechas acerca de la posible existencia de diferencias no observables. El hecho de que las personas con y sin seguro médico difieran en muchos aspectos visibles sugiere que incluso manteniendo fijas las características observadas, es probable que los no asegurados difieran de los asegurados en cuanto a aspectos que no vemos (al fin y al cabo, la lista de variables que podemos ver depende en parte del azar). En otras palabras, hasta en una muestra formada por



*Para salir del atolladero, recurre al azar*

Mi médico me dio seis meses de vida... pero cuando vio que no podía pagar la factura me dio seis meses más.

Walter Matthau

gente asegurada y no asegurada con el mismo nivel de formación, los mismos ingresos y la misma categoría laboral, los asegurados podrían tener valores más altos de $Y_{0i}$. El mayor desafío al que se enfrentan los maestros de la econometría consiste en eliminar el sesgo de selección que surge de esas diferencias no observadas.

La asignación aleatoria elimina el sesgo de selección. Puede que la logística de un experimento aleatorio sea compleja, pero la lógica es simple. Para estudiar los efectos de los seguros médicos en un experimento aleatorio, partiremos de una muestra de personas que no estén aseguradas en la actualidad. Después dotaremos de seguro médico a un subconjunto de esa muestra elegido al azar, y dejaremos que el resto recurra a los servicios de urgencias en caso de necesidad. Más tarde podrá cotejarse la salud del grupo asegurado y del no asegurado. La asignación aleatoria convierte en *ceteris paribus* esa comparación, puesto que los grupos de asegurados y de no asegurados por asignación aleatoria difieren tan sólo en si están o no asegurados y en cualquier consecuencia derivada de ello.

Supongamos que el Servicio de Salud del MIT decide no cobrar y lanza una moneda al aire para determinar la situación médica de los nuevos alumnos Ashish y Zandile (sólo esta vez, en deferencia a su distinguido Departamento de Economía). Zandile quedará asegurada si sale cruz, y en caso contrario la cobertura médica será para Ashish. Es un buen comienzo, pero no lo bastante, porque la asignación aleatoria a dos sujetos experimentales no produce manzanas aseguradas y no aseguradas. En primer lugar, Ashish es un chico y Zandile es una chica y, por regla general, las mujeres tienen mejor salud que los hombres. Si Zandile acabara teniendo mejor salud podría deberse a la suerte de haber nacido chica, sin ninguna relación con la suerte de haber ganado el sorteo del seguro médico. El problema aquí es que dos no bastan cuando se trata de usar una asignación aleatoria. Hay que asignar el tratamiento al azar dentro de una muestra lo bastante grande como para garantizar que se diluyan las diferencias debidas a características individuales, como el sexo.

Dos grupos elegidos al azar son comparables si son lo bastante amplios. Esto se debe a una propiedad estadística con un gran potencial que se conoce como *ley de los grandes números* (LGN). Esta ley caracteriza el comportamiento de las medias muestrales en relación con el tamaño de la muestra. En concreto, la ley de los grandes números

dice que una media muestral se puede acercar tanto como queramos a la media verdadera de la población analizada (por ejemplo, la población de estudiantes universitarios estadounidenses) simplemente ampliando el tamaño de la muestra.

Para ver esta ley en acción, juegue a los dados.[8] Es decir, lance un dado una vez y anote el resultado. Vuelva a lanzarlo y calcule la media de ambos resultados. Siga lanzando el dado y sacando medias. Existen las mismas probabilidades de que salga cualquier número del 1 al 6 (si el dado es perfecto), así que podemos esperar ver cada valor una cantidad igual de veces si practicamos este juego durante el tiempo suficiente. Como en este caso hay seis posibilidades y todas son igual de probables, el resultado esperado será una media con la misma ponderación para cada posibilidad, con todos los pesos iguales a 1/6:

$$\left(1 \times \frac{1}{6}\right) + \left(2 \times \frac{1}{6}\right) + \left(3 \times \frac{1}{6}\right) + \left(4 \times \frac{1}{6}\right) + \left(5 \times \frac{1}{6}\right) + \left(6 \times \frac{1}{6}\right)$$
$$= \frac{1 + 2 + 3 + 4 + 5 + 6}{6} = 3{,}5.$$

Este valor medio de 3,5 se denomina *valor esperado*, y en este caso se corresponde con el valor medio que se obtiene tras una cantidad infinita de lanzamientos de dados perfectos. El concepto de valor esperado es importante en nuestro trabajo, de modo que pasamos a dar su definición formal.

Valor esperado El valor esperado (o esperanza matemática) de una variable, $Y_i$, escrita $E[Y_i]$, es la media poblacional de esa variable. Si $Y_i$ es una variable obtenida a partir de un proceso aleatorio, como el lanzamiento de un dado, $E[Y_i]$ sería la media de infinitas repeticiones de dicho proceso. Si $Y_i$ es una variable procedente de una encuesta muestral, $E[Y_i]$ sería la media que resultaría si se encuestara a todas las personas de la población de la que se extrajo la muestra.

[8] Cubos de seis caras que en cada una de ellas portan desde uno hasta seis puntos. Los teléfonos inteligentes incluyen una aplicación que los simula.

Si lanzamos un dado pocas veces, la media del resultado tal vez diste mucho del valor esperado correspondiente. Si lo lanzamos dos veces, por ejemplo, y salen dos seises o dos unos, la media daría un resultado muy alejado del valor esperado de 3,5. Pero a medida que aumente el número de lanzamientos, la media tenderá con seguridad a 3,5. Es la ley de los grandes números en acción (y es el método de los casinos para obtener beneficios: en la mayoría de los juegos de azar no se puede ganar a la banca a la larga, porque la ganancia esperada de los jugadores es negativa). Pero lo más llamativo es que no hacen falta demasiados lanzamientos, ni una muestra demasiado grande, para que una media muestral se acerque al valor esperado. El apéndice del capítulo aborda de qué manera el número de lanzamientos o el tamaño de una encuesta muestral condiciona la precisión estadística.

Cuando se trata de experimentos aleatorios se generan muestras experimentales tomando subconjuntos de la población que se quiere estudiar, en lugar de repetir un juego varias veces, pero la ley de los grandes números funciona exactamente igual. Los sujetos de la muestra se reparten de forma aleatoria (como si se lanzara una moneda al aire) entre los grupos de tratamiento y de control, y se extraen de la misma población de partida. Así que la ley de los grandes números garantiza que los individuos de las muestras de tratamiento y de control asignados de forma aleatoria serán similares, siempre que las muestras sean lo bastante amplias. Por ejemplo, esperamos encontrar proporciones parecidas de hombres y mujeres en grupos de control y de tratamiento asignados al azar. La asignación aleatoria también genera grupos muy semejantes en cuanto a edad y nivel de formación. De hecho, los grupos creados mediante asignación aleatoria deberían parecerse en todo, incluso en aspectos que no se pueden medir ni apreciar con facilidad. Esta es la base del formidable poder de la asignación aleatoria para eliminar el sesgo de selección.

El poder de la asignación aleatoria se puede describir con precisión usando la definición siguiente, que está íntimamente relacionada con la definición de valor esperado (o esperanza matemática).

Valor esperado condicionado El valor esperado condicionado (o esperanza condicionada) de una variable, $Y_i$, dada una variable binaria, $D_i = 1$, se escribe $E[Y_i|D_i = 1]$. Esta es la media de $Y_i$ den-

tro de la población que tiene $D_i = 1$. De manera análoga, el valor esperado condicionado de una variable, $Y_i$, dada $D_i = 0$, escrita $E[Y_i|D_i = 0]$, es la media de $Y_i$ dentro de la población que tiene $D_i = 0$. Si $Y_i$ y $D_i$ son variables generadas mediante un proceso aleatorio, como el lanzamiento de un dado en diferentes circunstancias, entonces $E[Y_i|D_i = d]$ sería igual a la media de una cantidad infinita de repeticiones de ese proceso, siempre que las circunstancias indicadas por $D_i$ se mantengan fijas en $d$. Si $Y_i$ y $D_i$ proceden de una encuesta muestral, entonces $E[Y_i|D_i = d]$ es la media calculada cuando la muestra incluye todos los individuos de la población que tienen $D_i = d$.

Como los grupos de tratamiento y de control con asignación aleatoria proceden de la misma población de partida, son iguales en todo, incluso en su $Y_{0i}$ esperado. En otras palabras, los valores esperados condicionados, $E[Y_i|D_i = 1]$ y $E[Y_i|D_i = 0]$, son iguales. Y esto significa a su vez que:

La asignación aleatoria elimina el sesgo de selección Cuando $D_i$ se asigna aleatoriamente, $E[Y_{0i}|D_i = 1] = E[Y_{0i}|D_i = 0]$, y la diferencia de valores esperados debida a pertenecer o no al grupo de tratamiento se corresponde con el efecto causal del tratamiento:

$$\begin{aligned} &E[Y_i|D_i = 1] - E[Y_i|D_i = 0] \\ &\quad = E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 0] \\ &\quad = E[Y_{0i} + \kappa|D_i = 1] - E[Y_{0i}|D_i = 0] \\ &\quad = \kappa + E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0] \\ &\quad = \kappa. \end{aligned}$$

Suponiendo que la muestra en cuestión sea lo bastante grande como para que la ley de los grandes números obre su magia (de forma que se puedan reemplazar las medias condicionadas en la ecuación (1.4) por los valores esperados condicionados), el sesgo de selección desaparece de un experimento aleatorio. La asignación aleatoria funciona, no porque elimine las diferencias individuales, sino porque garantiza que la mezcla de los individuos que se están comparando es la misma. Pensemos en toneles llenos de manzanas y naranjas en la misma proporción. Tal como se explica en los si-

guientes capítulos, la aleatorización no es la única vía para generar estas comparaciones *ceteris paribus*, pero la mayoría de los maestros la consideran la mejor.

A la hora de analizar datos procedentes de un experimento aleatorio o cualquier otro diseño de investigación, los maestros casi siempre empiezan comprobando si los grupos de tratamiento y de control resultan ser similares. Este proceso, llamado *comprobación del equilibrio*, equivale a una comparación de medias muestrales como las del apartado B de la tabla 1.1. Las características medias del apartado B resultan ser diferentes o desequilibradas, lo que indica el hecho de que los datos de esta tabla no proceden de nada parecido a un experimento aleatorio. Vale la pena comprobar el equilibrio de este modo cada vez que haya que calcular efectos causales.

La asignación aleatoria de seguros médicos parece una propuesta descabellada. Pero resulta que la cobertura con un seguro médico se ha asignado aleatoriamente dos veces a amplias muestras representativas de los estadounidenses. El RAND Health Insurance Experiment (HIE) [Experimento de Seguros Médicos RAND], que se desarrolló desde 1974 hasta 1982, fue uno de los experimentos sociales más influyentes de la historia de la investigación. El HIE contó con 3.958 personas de edades comprendidas entre 14 y 61 años de seis zonas de Estados Unidos. La muestra HIE excluía a los beneficiarios de Medicare y a la mayoría de los suscritos a Medicaid o a seguros médicos militares. A los participantes en el HIE se les practicó una asignación aleatoria de un plan de seguros entre 14 posibles. Los participantes no tuvieron que pagar primas por el seguro médico, pero cada plan tenía cierta variedad de prestaciones relacionadas con copagos, lo cual implicaba grandes diferencias en cuanto a la cobertura ofrecida.

El plan de seguros más generoso del HIE daba asistencia integral sin ningún copago. En el extremo opuesto del espectro figuraban tres planes que cubrían situaciones catastróficas que exigían a las familias el pago del 95% de sus gastos médicos, aunque esos gastos estaban limitados en proporción a los ingresos (o limitados a 1.000 dólares por familia, si esta cantidad era menor). Los planes de cobertura en situaciones catastróficas eran casi como no tener seguro. Un segundo tipo de seguro médico (el plan «con franquicia individual») también obligaba a las familias a pagar el 95% de los gastos por asistencia am-

bulatoria, pero con un límite de 150 dólares por persona o 450 dólares por familia. Un conjunto de otros nueve planes ofrecía cierta variedad de provisiones de seguro que exigían que los participantes cubrieran desde el 25% hasta el 50% de los gastos, pero siempre limitados a una proporción de sus ingresos o a 1.000 dólares, a la cantidad que fuera menor de las dos. Las familias participantes se apuntaron a los planes del experimento durante un intervalo de entre tres y cinco años y aceptaron renunciar a cualquier cobertura de seguros previa, a cambio de recibir una cantidad fija al mes independiente del uso que hicieran de la asistencia sanitaria.[9]

El HIE se debió sobre todo a un interés en lo que los economistas denominan la elasticidad precio de la demanda de servicios de asistencia médica. En concreto, los investigadores del RAND querían averiguar si cae el uso de los servicios de asistencia médica, y en qué medida, cuando aumenta el precio de la atención sanitaria. Las familias pertenecientes al plan de asistencia gratuita tenían un coste cero, mientras que los seguros con copago reducían los gastos realizados en un 25 o un 50%, y las familias que estaban cubiertas en situaciones catastróficas y con planes que les obligaban a pagar casi la totalidad del coste de los servicios, al menos hasta alcanzar el máximo de gasto establecido. Pero los investigadores también querían saber si los seguros médicos con coberturas más integrales y generosas deparan mejor salud. La respuesta a la primera cuestión estaba clara, era un «sí»: el uso de los servicios sanitarios depende enormemente del precio de la atención. La respuesta al segundo interrogante es más turbia.

### *Resultados aleatorios*

Los experimentos de campo aleatorios son más intrincados que el simple lanzamiento de una moneda al aire, a veces por desgracia. El HIE fue complicado por tener muchos grupos de tratamiento con un

[9] Nuestra descripción del HIE se basa en el artículo de Robert H. Brook *et al.* titulado «Does Free Care Improve Adults' Health? Results from a Randomized Controlled Trial», *New England Journal of Medicin*e, vol. 309, n.º 23, 8 de diciembre de 1983, páginas 1.426-1.434. Para consultar un análisis reciente, véase además Aviva Aron-Dine, Liran Einav y Amy Finkelstein, «The RAND Health Insurance Experiment, Three Decades Later», *Journal of Economic Perspectives*, vol. 27, invierno de 2013, páginas 197-222.

tamaño pequeño, repartidos en más de una docena de planes de seguros. Al ser demasiado pequeños los grupos de tratamiento asociados a cada plan, las comparaciones entre ellos carecen de significación estadística. De ahí que casi todos los análisis de los datos del HIE comiencen agrupando sujetos de acuerdo con la similitud entre los planes del HIE que les fueron asignados. Nosotros también lo haremos aquí.[10]

Una manera natural de agrupar los planes consiste en combinarlos de acuerdo con la cantidad de gastos compartidos que exigen. Los tres planes que dan cobertura a situaciones catastróficas, cuyos suscriptores asumen casi todos sus gastos médicos hasta un límite bastante alto, vienen a ser como no tener ningún seguro. El plan con franquicia individual brindaba una cobertura mayor, pero sólo reduciendo el límite de gastos totales que los sujetos a este plan estaban obligados a afrontar. Los nueve planes de seguro con copago ofrecían una cobertura más sustanciosa al repartir los gastos de atención médica de los suscriptores con la aseguradora desde el primer dólar gastado. Por último, el plan gratuito consistía en una actuación extrema que era de esperar que generara el mayor aumento en el uso de la asistencia sanitaria y, tal vez, también en la salud. Esta clasificación nos permite distinguir hasta cuatro grupos de planes: catastróficos, con franquicia, de copago y gratuitos, en lugar de los 14 planes iniciales. Los planes de cobertura en situaciones catastróficas proporcionan el

[10] Otras complicaciones del HIE incluyen el hecho de que en lugar de lanzar una moneda al aire (o hacer algo equivalente mediante computadora), los investigadores del RAND aplicaron un esquema complejo de asignaciones que puede repercutir en las propiedades estadísticas de los análisis resultantes (para ver los detalles, consúltese Carl Morris, «A Finite Selection Model for Experimental Design of the Health Insurance Study», *Journal of Econometrics*, vol. 11, n.º 1, septiembre de 1979, páginas 43-61). Las intenciones eran buenas, porque los que realizaron el experimento esperaban evitar así la desviación por azar del equilibrio perfecto entre los grupos de tratamiento. La mayoría de analistas del HIE ignoran las complicaciones estadísticas resultantes, aunque es probable que muchos coincidan con nosotros en lamentar este intento de rizar el rizo de la asignación aleatoria. Un problema más serio radica en el gran número de individuos del HIE que abandonaron el experimento, y en las grandes diferencias en cuanto a índices de atrición en cada grupo de tratamiento (por ejemplo, el plan gratuito lo abandonó menos gente). Tal como señalaron Aron-Dine, Einav y Finkelstein, «The RAND Experiment», *Journal of Economic Perspectives*, 2013, las diferencias en cuanto a abandono han podido poner en peligro la validez del experimento. Los «randomistas» actuales mejoran los problemas debidos a estos detalles de diseño (véanse, por ejemplo, los experimentos descritos en la obra *Repensar la pobreza: un giro radical en la lucha contra la desigualdad global*, de Abhijit Banerjee y Esther Duflo, Madrid: Taurus, 2011, trad. de F. Javier Mato Díaz).

mismo control (aproximado) que no tener ningún seguro, mientras que los planes con franquicia, de copago y gratuitos se caracterizan por tener niveles cada vez más elevados de cobertura.

Al igual que en las comparaciones no experimentales, un primer paso en nuestro análisis experimental consiste en comprobar el equilibrio. ¿Se parecen los sujetos asignados a los grupos de tratamiento y de control (en este caso, a esquemas de seguros médicos que van desde una cobertura mínima hasta una cobertura total)? Lo mediremos comparando características demográficas y datos médicos recopilados antes de que comenzara el experimento. Como las características demográficas no cambian, mientras que las variables de salud en cuestión se midieron antes de la asignación aleatoria, esperamos encontrar tan sólo pequeñas diferencias con respecto a esas variables en los grupos asignados a distintos planes.

En contraste con la comparación que establecimos entre las características de los que respondieron la Encuesta Nacional de Salud (NHIS) en relación con su situación en cuanto a seguro médico en la tabla 1.1, una comparación de las características entre grupos de tratamiento de asignación aleatoria en el experimento del RAND revela una gran similitud entre la gente asignada a distintos planes del HIE. Esto se ve en el apartado A de la tabla 1.3. La columna (1) de esta tabla da medias del grupo del plan catastrófico, mientras que el resto de columnas comparan los grupos a los que se les asignaron coberturas de seguros más generosas con el grupo catastrófico de control. Como medida global, la columna (5) compara una muestra que combina sujetos de los planes con franquicia, de copago y gratuitos con sujetos de los planes catastróficos. Los individuos asignados a planes con coberturas más generosas tienen una probabilidad algo menor de ser mujeres y de tener un nivel de formación algo inferior que los de los planes catastróficos. Asimismo, se aprecia alguna variación en cuanto a ingresos, pero las diferencias entre los grupos de cada plan son en su mayoría pequeñas y tienen la misma probabilidad de tender hacia un lado que hacia el otro. Este patrón contrasta con las grandes y sistemáticas diferencias demográficas entre asegurados y no asegurados que se aprecian en los datos del NHIS resumidos en la tabla 1.1.

Parece probable que las pequeñas diferencias entre grupos que se aprecian en el apartado A de la tabla 1.3 reflejen una variación aleatoria surgida de manera natural como parte del proceso de muestreo.

En cualquier muestra estadística, las diferencias aleatorias aparecen porque analizamos una de entre muchas extracciones posibles a partir de la población de la que procede la muestra. Cabría esperar que

**Tabla 1.3. Características demográficas y salud de base en el estudio RAND HIE**

| | **Medias** | **Diferencias entre grupos de plan de salud** | | | |
|---|---|---|---|---|---|
| | Plan catastrófico | Deducible catastrófico | Coseguro catastrófico | Gratuito catastrófico | Seguro de cualquier tipo catastrófico |
| | (1) | (2) | (3) | (4) | (5) |
| **A. Características demográficas** | | | | | |
| Mujeres | .560 | –.023 | –.025 | –.038 | –.030 |
| | | (.016) | (.015) | (.015) | (.013) |
| No blancos | .172 | –.019 | –.027 | –.028 | –.025 |
| Edad | 32.4 | .56 | .97 | .43 | .64 |
| | [12.9] | (.68) | (.65) | (.61) | (.54) |
| Formación académica | 12.1 | –.16 | | | |
| | [2.9] | (.19) | (.19) | (.18) | (.16) |
| Ingresos familiares | 31,603 | –2,104 | 970 | | |
| | [18,148] | (1,384) | (1,389) | (1,345) | (1,181) |
| Hospitalizados el año anterior | .115 | .004 | –.002 | .001 | .001 |
| | | (.016) | (.015) | (.015) | (.013) |
| **B. Variables sanitarias de base** | | | | | |
| Índice de salud general | 70.9 | –1.44 | .21 | –1.31 | –.93 |
| | [14.9] | (.95) | (.92) | (.87) | (.77) |
| Colesterol (mg dl) | 207 | –1.42 | –1.93 | –5.25 | –3.19 |
| | [40] | (2.99) | (2.76) | (2.70) | (2.29) |
| Presión sanguínea sistólica (mm Hg) | 122 | 2.32 | .91 | 1.12 | 1.39 |
| | [17] | (1.15) | (1.08) | (1.01) | (.90) |
| Índice de salud mental | 73.8 | –.12 | 1.19 | .89 | .71 |
| | [14.3] | (.82) | (.81) | (.77) | (.68) |
| Tamaño de la muestra | 759 | 881 | 1,022 | 1,295 | 3,198 |

*Notas:* Esta tabla describe las características demográficas y la salud de base de los sujetos en el experimento RAND HIE (Health Insurance Experiment, «Experimento sobre Seguros Sanitarios»). La columna (1) muestra la media para el grupo al que se asignó un seguro catastrófico. Las columnas (2) a (5) comparan con el grupo anterior las medias para los grupos con seguros de tipo deducible, en régimen de copago, con prestaciones gratuitas, o todos juntos. Los errores típicos se dan entre paréntesis para las columnas (2) a (5). Para la columna (1) constan las desviaciones típicas entre corchetes.

otra muestra de un tamaño similar extraída de la misma población condujera a comparaciones parecidas (aunque no idénticas) a las de la tabla. Las herramientas de la inferencia estadística permiten valorar qué nivel de variación cabría esperar entre una muestra y otra.

El apéndice de este capítulo ofrece una explicación breve de cómo cuantificar la variación muestral por medio de pruebas estadísticas formales. Tales pruebas recurren a la yuxtaposición de diferencias entre las medias de grupos por medio de sus *errores típicos*, los números que figuran entre paréntesis bajo las diferencias de valores medios en las columnas (2) a (5) de la tabla 1.3. El error típico de una diferencia de medias ofrece una medida de su precisión estadística: cuando una diferencia entre medias de grupos es menor que unas dos veces los errores típicos, entonces se suele considerar que esa diferencia puede deberse al azar, y que su aparición es compatible con la hipótesis de que las poblaciones de las que se han extraído las muestras son, en realidad, la misma.

Si las diferencias resultan superiores a unas dos veces los errores típicos, entonces se califican como *estadísticamente significativas*: en estos casos es muy improbable que la aparición de tales diferencias se deba tan sólo al azar. Una diferencia que no sea estadísticamente significativa probablemente se deba a las irregularidades del proceso de elección de las muestras. La noción de significancia estadística nos ayuda a interpretar comparaciones como las de la tabla 1.3. Las diferencias que figuran en esta tabla son pequeñas en su mayoría, con sólo dos de ellas (la proporción de mujeres en las columnas [4] y [5]) por encima del doble de los correspondientes errores típicos. También cabe atribuir al azar la aparición de unas pocas diferencias aisladas que son estadísticamente significativas en tablas que contienen tantas comparaciones como la tabla 1.3. Reconforta comprobar, además, que los errores típicos de esta tabla no son muy grandes, lo que indica que las diferencias reales entre grupos se han medido con una precisión razonable.

El apartado B de la tabla 1.3 complementa los contrastes del apartado A con indicios razonables de que existe un buen equilibrio en los *resultados pretratamiento* entre los distintos grupos. Este apartado no exhibe diferencias estadísticas significativas en el índice pretratamiento referido a salud general. De igual modo parece claro que los valores pretratamiento de colesterol, presión sanguínea y salud men-

tal no guardan relación con el tratamiento asignado, con sólo un par de valores cercanos a la significancia estadística. Además, aunque el valor algo más bajo de los niveles de colesterol pueda sugerir una salud de partida un poco mejor en el grupo catastrófico, las diferencias en el índice de salud general entre estos dos grupos apuntan en el sentido contrario (dado que números inferiores señalan peor salud). La ausencia de un patrón sistemático refuerza la idea de que estas discrepancias se deben al azar.

El primer hallazgo relevante que surge del HIE es que los sujetos asignados a planes de seguros más generosos recurrieron mucho más a la atención sanitaria. Esta conclusión, que refuerza la idea de los economistas de que la demanda de un bien debería aumentar si éste se vuelve más barato, se deduce del apartado A de la tabla 1.4.[11] Como era de esperar, los ingresos hospitalarios resultaron ser menos sensibles al precio que la atención extrahospitalaria, probablemente porque la decisión de efectuar o no un ingreso hospitalario suele tomarla el personal médico. Por otra parte, la asignación al plan de salud con atención gratuita hizo aumentar el gasto por paciente en dos tercios (169.248) si se compara con el gasto de los planes catastróficos, mientras que los gastos médicos totales aumentan en un 45%. Estas diferencias tan grandes son relevantes desde el punto de vista económico y, a la vez, son estadísticamente significativas.

Los sujetos que no tenían que preocuparse por el coste de la atención sanitaria muestran una tendencia clara a usarla bastante más. Pero esta atención adicional, ¿mejoró la salud de estas personas? El apartado B de la tabla 1.4 compara los indicadores de salud entre los grupos de tratamiento del HIE, y parece indicar que la respuesta es negativa. Los niveles de colesterol, de presión sanguínea y los índices globales de salud general y de salud mental son muy similares entre todos los grupos (los resultados se midieron sobre todo al cabo de tres o cinco años tras la asignación aleatoria). Las pruebas estadísticas formales no arrojan diferencias estadísticamente significativas, como ponen de manifiesto tanto las comparaciones entre grupos

[11] Los resultados RAND que aportamos se basan en nuestra propia tabulación del archivo de uso público del estudio HIE, tal como se describe en el apartado de «Notas empíricas» al final de este libro. Los resultados RAND originales se resumen en la obra de Joseph P. Newhouse *et al., Free for All? Lessons from the RAND Health Insurance Experiment*, Harvard University Press, 1994.

concretos (columnas (2) a (4)) como las diferencias de salud entre los sujetos del grupo del plan catastrófico y el conjunto de quienes se asignaron a planes más generosos (columna [5]).

**Tabla 1.4. Gasto sanitario y resultados de salud del estudio RAND HIE**

| | Medias | Diferencias entre grupos de plan de salud | | | |
|---|---|---|---|---|---|
| | Plan catastrófico | Deducible catastrófico | Coseguro catastrófico | Gratuito catastrófico | Seguro de cualquier tipo catastrófico |
| | (1) | (2) | (3) | (4) | (5) |
| **A. Uso de los servicios sanitarios** | | | | | |
| Consulta médica | 2.78 | .19 | .48 | 1.66 | .90 |
| | [5.50] | (.25) | (.24) | (.25) | (.20) |
| Gastos en atención extrahospitalaria | 248 | 42 | 60 | 169 | 101 |
| | [488] | (21) | (21) | (20) | (17) |
| Ingresos hospitalarios | .099 | .016 | .002 | 029 | 017 |
| | [.379] | (.011) | (.011) | (.010) | (.009) |
| Gastos en atención hospitalaria | 388 | 72 | 93 | 116 | 97 |
| | [2,308] | (69) | (73) | (60) | (53) |
| Gasto total | 636 | 114 | 152 | 285 | 198 |
| | [2,535 | (79) | (85) | (72) | (63) |
| **B. Resultados de salud** | | | | | |
| Índice general de salud | 68.5 | –.87 | .61 | –.78 | –.36 |
| | [15.9] | (.69) | (.90) | (.87) | (.77) |
| Colesterol (mg dl) | 203 | .69 | –2.31 | –1.83 | –1.32 |
| Presión sanguínea sistólica | 122 | 1.17 | –1.39 | –.52 | –.36 |
| (mm Hg) | [19] | (1.06) | (.99) | (.93) | (.85) |
| Índice de salud mental | 75.5 | .45 | 1.07 | .43 | .64 |
| | [14.8] | (.91) | (.87) | (.83) | (.75) |
| Tamaño de la muestra | 759 | 881 | 1,022 | 1,295 | 3,198 |

*Notas:* Esta tabla incluye las medias y los efectos del tratamiento en gastos sanitarios y repercusión en la salud del experimento RAND HIE (Health Insurance Experiment, «Experimento sobre Seguros Sanitarios»). La columna (1) muestra la media para el grupo al que se asignó un seguro catastrófico. Las columnas (2) a (5) comparan con el grupo anterior las medias para los grupos con seguros de tipo deducible, en régimen de copago, con prestaciones gratuitas, o todos juntos. Los errores típicos se dan entre paréntesis para las columnas (2) a (5). Para la columna (1) constan las desviaciones típicas entre corchetes.

Estos resultados del HIE convencieron a muchos economistas de que los planes de salud generosos pueden tener consecuencias indeseadas y no buscadas, puesto que aumentan el uso de los servicios sanitarios y sus costes sin proporcionar resultados en términos de mejora de la salud.

## 1.2 El rastro de Oregón

Maestro Kan: Es difícil comprender la verdad.
Kwai Chang Caine: Pero esto es un hecho, no una verdad. La verdad suele permsanecer oculta, como una sombra en la oscuridad.

*Kung Fu,* primera temporada, episodio 14

El experimento HIE supuso un intento ambicioso de valorar el impacto de los seguros médicos sobre los costes de los servicios sanitarios y sobre la salud. A pesar de ello, a juzgar por el curso que siguen los debates actuales sobre los seguros médicos, es posible que el HIE no diera en el blanco. Por un motivo: todos los grupos de tratamiento del HIE incluían al menos cobertura en casos catastróficos, de manera que la deuda contraída en concepto de gastos sanitarios estaba limitada en todos los casos. Más importante aún resulta el hecho de que los estadounidenses no asegurados de hoy día difieren mucho de la población incluida en el experimento HIE: la mayoría de los no asegurados son más jóvenes, tienen menos estudios, son más pobres y es menos probable que tengan trabajo. El valor de una atención médica adicional para un grupo así podría ser muy distinto del que tiene para las familias de clase media que participaron en el HIE.

Una de las ideas más controvertidas en el escenario actual de las políticas de salud en Estados Unidos corresponde a la extensión de Medicaid hasta cubrir a las personas que ahora no están aseguradas (llama la atención que cuando se concibió el experimento RAND de lo que se hablaba era de ampliar Medicare, el seguro médico público para personas mayores). En este momento Medicaid cubre a las familias que reciben subsidios del Estado, a algunos discapacitados, a la infancia pobre y a las mujeres embarazadas sin recursos. Supongamos que pretendemos extender Medicaid para que incluya también a quienes no están incluidos en estas categorías. ¿Cómo afectaría esta

ampliación a los gastos en atención sanitaria? ¿Tendería a desplazar los tratamientos desde los caros y masificados servicios de urgencias hacia una atención primaria quizá más efectiva? ¿Mejoraría la salud con la ampliación de Medicaid?

Muchos estados de la Unión han empezado a «experimentar» con la extensión de Medicaid en el sentido de ampliar los criterios de inclusión, cargando la mayor parte de la factura al gobierno federal. Pero tengamos en cuenta que estas iniciativas no constituyen experimentos verdaderos, porque toda persona que cumpla los requisitos ampliados para entrar en Medicaid pasa a obtener esa cobertura. El modo más convincente de conocer las consecuencias de la extensión de Medicaid consistiría en ofrecer la cobertura de manera aleatoria a personas que ahora pertenecen a grupos no elegibles. Pero tal vez sea pedir demasiado. Aun así, en un experimento social admirable, el estado de Oregón ha ofrecido recientemente Medicaid a miles de personas elegidas al azar, en un sorteo de seguros sanitarios anunciado de manera pública.

Podría parecernos que el sorteo de seguros médicos de Oregón consiste en una selección aleatoria de ganadores y perdedores a partir de una lista de candidatos, pero el resultado del sorteo no es automático, ni siquiera para los ganadores. Lo que ganan las personas elegidas es la oportunidad de cursar su solicitud en el Plan de Salud de Oregón (OHP, Oregon Health Plan), la versión oregoniana de Medicaid. A continuación, el estado de Oregón revisa estas solicitudes y concede la cobertura sanitaria a las personas residentes en su territorio que tengan la ciudadanía estadounidense, o la condición de inmigrante legal, con edades entre diecinueve y sesenta y cuatro años, que no sean elegibles ya de partida para Medicaid, que lleven al menos seis meses sin seguro médico, con ingresos inferiores al umbral federal de pobreza, y que tengan poco patrimonio. Quienes ganan el sorteo tienen que documentar su nivel de pobreza y cumplimentar el papeleo en 45 días, para disponer de la cobertura de manera efectiva.

El sorteo OHP de 2008 responde más a razones de justicia que de investigación, pero eso no lo hace menos extraordinario, porque el sorteo oregoniano de salud nos ofrece algunos de los mejores indicios que podríamos desear acerca de los costes y los beneficios de la cobertura sanitaria para las personas que no disponen de ella, una

circunstancia que movió a la investigadora del MIT Amy Finkelstein, y sus colaboradores, a acometer un estudio sobre el OHP.[12]

Unas 75.000 personas se inscribieron en el sorteo para la ampliación de la atención médica del OHP. Unas 30.000 de ellas fueron elegidas al azar e invitadas a presentar su solicitud para el OHP: este es el grupo de tratamiento en este experimento, mientras que los otros 45.000 individuos constituyen la muestra de control.

La primera pregunta que surge en este contexto se refiere a si de verdad el hecho de ganar el sorteo OHP incrementa las probabilidades de conseguir el seguro médico. Está justificado plantearse esta incógnita porque algunos solicitantes ya cumplían de partida los requisitos originales para entrar en Medicaid con independencia del sorteo. El apartado A de la tabla 1.5 muestra que en torno a un 14% de los sujetos de control (perdedores en el sorteo) lograron la cobertura de Medicaid en el año posterior al primer sorteo. Al mismo tiempo, la columna segunda, que refleja las diferencias entre los grupos de tratamiento y de control, revela que la probabilidad de lograr el ingreso en Medicaid se incrementó en un 26% para los sujetos ganadores del sorteo. La columna (4) pone de manifiesto un incremento similar para la submuestra con domicilio en Portland o sus alrededores, la mayor ciudad de Oregón. Se concluye que quienes ganaron el sorteo lograron el seguro médico en una proporción mucho mayor que quienes lo perdieron, una diferencia que podría afectar a su utilización de los servicios médicos y a su salud.[13]

El grupo de tratamiento OHP (es decir, quienes ganaron en el sorteo) recurrió a los servicios de salud más de lo que lo habrían hecho en caso de no ganar. Esta circunstancia se aprecia en la tabla 1.5, que muestra estimaciones del cambio en el uso del servicio en las filas que

[12] Véase Amy Finkelstein *et al.*, «The Oregon Health Insurance Experiment: Evidence from the First Year», *Quarterly Journal of Economics*, vol. 127, número 3, agosto de 2012, páginas 1057-1106; Katherine Baicker *et al.*, «The Oregon Experiment-Effects of Medicaid on Clinical Outcomes», *New England Journal of Medicine*, vol. 368, número 18, 2 de mayo de 2013, páginas 1713-1722; y Sarah Taubman *et al.*, «Medicaid Increases Emergency Department Use: Evidence from Oregon's Health Insurance Experiment», *Science*, vol. 343, número 6.168, 17 de enero de 2014, páginas 263-268.

[13] ¿Por qué no lograron seguro médico todos los ganadores del sorteo? Algunos no consiguieron reunir todo el papeleo a tiempo, mientras que cerca de la mitad de quienes lo cumplimentaron dentro de plazo resultaron no ser elegibles en la revisión posterior.

**Tabla 1.5. Efectos del OHP sobre la cobertura sanitaria y el uso de servicios médicos**

| | Oregón | | Área de Portland | |
|---|---|---|---|---|
| | Media de control | Efecto del tratamiento | Media de control | Efecto del tratamiento |
| Resultado | (1) | (2) | (3) | (4) |
| **A. Datos administrativos** | | | | |
| Estuvieron antes en Medicaid | .141 | .256 | .151 | .247 |
| | | (.004) | | (.006) |
| Tuvieron antes ingresos hospitalarios | .067 | .005 | | |
| | | (.002) | | |
| Visitaron antes los servicios de urgencias | | | .345 | .017 |
| | | | | (.006) |
| Número de visitas a los servicios de urgencias | | | 1.02 | .101 |
| | | | | (.029) |
| Tamaño de la muestra | 74,922 | | 24,646 | |
| **B. Datos del estudio** | | | | |
| Consultas médicas (en los últimos seis meses) | 1.91 | .314 | | |
| | | (.054) | | |
| ¿Alguna receta? | .637 | .025 | | |
| | | (.008) | | |
| Tamaño de la muestra | 23,741 | | | |

*Notas:* Esta tabla incluye la estimación de los efectos de haber ganado el sorteo del Plan de Salud de Oregón (OHP, Oregon Health Plan) sobre la cobertura sanitaria y sobre la utilización de los servicios médicos. Las columnas impares muestran las medias para el grupo de control. Las columnas pares recogen el coeficiente de la regresión efectuada por medio de una variable binaria sólo sobre los sujetos ganadores del sorteo. Los errores típicos figuran entre paréntesis.

aparecen por debajo de las estimaciones del efecto OHP en la cobertura de Medicaid. El índice de hospitalización aumentó cerca de medio punto porcentual, un efecto modesto, aunque estadísticamente significativo. La visita a las urgencias médicas y a consultas externas, así como los medicamentos recetados experimentan todos un incre-

mento destacado. El hecho de que el número de visitas a urgencias suba alrededor de un 10%, un efecto estimado con precisión (el error típico asociado a este estimado consta en la columna (4) y asciende a 0,029) es especialmente digno de mención. Muchos responsables políticos esperaban que, al implementar los seguros sanitarios, los pacientes que antes no estaban asegurados dejaran de acudir a urgencias y usaran otros servicios sanitarios menos caros.

Finalmente, el resultado de la historia de los seguros médicos aparece en la tabla 1.6: la muestra de personas que ganaron el sorteo presenta un incremento modesto en la probabilidad de que consideren que gozan de una salud buena o mejor (un efecto que asciende a 0,039 puntos y que puede compararse con la media del grupo de control, que es de 0,55; la variable «Buena salud» es binaria). Los resultados de entrevistas personalizadas efectuadas en Portland parecen indicar que este incremento procede más de una mejora en salud mental que de una ganancia en salud física, tal como se comprueba en las filas segunda y tercera de la columna (4) (las variables de salud en la muestra de Portland consisten en índices que adoptan valores entre 0 y 100). Como en el experimento RAND, los resultados de Portland sugieren que los indicadores de salud física, como el colesterol o la presión sanguínea, no se ven afectados en general por el acceso al seguro médico OHP.

El hecho de que el sorteo OHP tuviera unos efectos tan leves sobre la salud decepcionó a los políticos que esperaban que un seguro proporcionado por el Estado generara una mejora sanitaria para los estadounidenses de ingresos bajos. El efecto de que el seguro médico incrementara, en vez de reducir, la utilización de los costosos servicios de urgencias resultó especialmente frustrante. A la vez, el apartado B de la tabla 1.6 revela que el seguro médico proporcionó la seguridad financiera para la que se diseñó. En particular, resulta menos probable que las unidades familiares que ganaron el sorteo incurran en gastos médicos elevados o que acumulen deudas por la necesidad de pagar cuidados médicos. Es posible que esta mejora en la salud financiera sea la responsable de la mejora en salud mental que se detecta en el grupo de tratamiento.

También vale la pena remarcar que los efectos sobre la salud financiera y mental que se aprecian en la tabla 1.6 proceden, probablemente, del 25% de individuos de la muestra que accedieron al seguro

**Tabla 1.6. Efectos del OHP sobre los indicadores de salud y sobre la salud financiera**

| | Oregón | | Área de Portland | |
|---|---|---|---|---|
| | Media de control | Efecto del tratamiento | Media de control | Efecto del tratamiento |
| Resultado | (1) | (2) | (3) | (4) |
| **A. Indicadores de salud** | | | | |
| Buena salud | .548 | .039 | | |
| | | (.008) | | |
| Índice de salud física | | | 45.5 | .29 |
| | | | | (.21) |
| Índice de salud mental | | | 44.4 | .47 |
| | | | | (.24) |
| Colesterol | | | 204 | .53 |
| | | | | (.69) |
| Presión sanguínea sistólica (mm Hg) | | | 119 | –.13 |
| | | | | (.030) |
| **B. Salud financiera** | | | | |
| Gastos médicos superiores al 30% de los ingresos | | | .055 | –.011 |
| | | | | (.005) |
| ¿Deudas por causas médicas? | | | .568 | –.032 |
| | | | | (.010) |
| Tamaño de la muestra | 23,741 | | 12,229 | |

*Notas:* Esta tabla incluye la estimación de los efectos de haber ganado el sorteo del Plan de Salud de Oregón (OHP, Oregon Health Plan) sobre los indicadores de salud y sobre la salud financiera. Las columnas impares muestran las medias para el grupo de control. Las columnas pares recogen el coeficiente de la regresión efectuada por medio de una variable binaria sólo sobre los sujetos ganadores del sorteo. Los errores típicos figuran entre paréntesis.

médico como consecuencia del sorteo. Si se tiene en cuenta el hecho de que muchos ganadores no cambiaron de seguro médico, se deduce que los incrementos en seguridad financiera y mental para ese cuarto de los solicitantes que obtuvieron el seguro como resultado del sorteo fueron considerablemente mayores de lo que resulta al comparar simplemente ganadores con perdedores. El capítulo 3, dedicado a los métodos de variables instrumentales, da los detalles sobre la naturaleza de tales ajustes. Como veremos pronto, el ajuste que hay que aplicar es

el resultado de dividir las diferencias de resultados ganadores-perdedores entre la diferencia ganadores-perdedores en probabilidad de tener seguro. Esto implica que el efecto de tener seguro médico es nada menos que cuatro veces mayor que el efecto de haber ganado el sorteo OHP (este ajuste no afecta a la significancia estadística).

Las conclusiones de los experimentos RAND y de Oregón son bastante similares. Dos iniciativas similares orientadas a poblaciones muy diferentes evidencian que la utilización de los servicios sanitarios se incrementa de manera considerable en respuesta a un aumento de la cobertura, mientras que ninguno de los dos experimentos revela efectos destacados del seguro médico sobre la salud física. En 2008, los ganadores del sorteo OHP disfrutaron de mejoras modestas pero apreciables en cuanto a salud mental. Es importante destacar que el experimento OHP tuvo también el éxito al proteger a muchos ganadores del sorteo de las consecuencias económicas de una mala salud, justo como debería hacerlo una buena política de aseguramiento. A la vez, estos estudios indican que no se debe esperar que los seguros médicos pagados con dinero público arrojen como resultado una mejora espectacular de la salud.

Maestro Joshway: Por favor, Pequeño Saltamontes, sé breve.

Pequeño saltamontes: La inferencia causal compara resultados potenciales, descripciones del mundo que resultaría cuando se eligen caminos alternativos.

Maestro Joshway: ¿Comparamos los caminos elegidos por unos con los caminos que tomaron otros?

Pequeño Saltamontes: Tales comparaciones suelen estar contaminadas por sesgos de selección, es decir, por diferencias entre sujetos tratados y sujetos de control que habrían existido incluso en ausencia de tratamiento.

Maestro Joshway: ¿Es posible eliminar el sesgo de selección?

Pequeño Saltamontes: La asignación aleatoria de tratamiento y de control eliminan el sesgo de selección. A pesar de ello, incluso en los experimentos aleatorios conviene comprobar que haya equilibrio.

Maestro Joshway: ¿Hay una única verdad causal que una investigación aleatoria tenga la seguridad de revelar?

Pequeño Saltamontes: Veo ahora que puede haber muchas verdades, Maestro, algunas compatibles, otras contradictorias. Por tanto, prestamos una atención especial a los resultados de dos o más experimentos de naturaleza similar.

## Maestros de la econometría: de Daniel a R. A. Fisher

Ya el Viejo Testamento recoge la importancia de los grupos de control. El libro de Daniel narra cómo el rey babilonio Nabucodonosor decidió preparar a Daniel y a otros cautivos israelitas para su servicio real. En el contexto de la esclavitud no fue tan mala cosa, porque el rey ordenó que sus sirvientes recibieran «una porción para cada día de la comida del rey y del vino que él bebía». Daniel no se sentía cómodo con tan rica dieta y prefería una alimentación vegetariana más modesta. Los cortesanos del rey denegaron en principio la solicitud de Daniel de recibir una comida especial por temor a que una dieta así resultara inadecuada para una persona dedicada al servicio del rey. Pero a Daniel no le faltó la audacia para proponer un experimento controlado: «Prueba, te ruego, con tus siervos *por* diez días, y que nos den legumbres para comer y agua para beber. Parezcan luego delante de ti nuestros rostros, y los rostros de los muchachos que comen de la porción de la comida del rey; y según lo que vieres, harás con tus siervos». (Daniel 1, 12-13). La Biblia relata cómo este experimento dio apoyo a la conjetura de Daniel respecto al carácter saludable, en términos relativos, de la dieta vegetariana, aunque por lo que sabemos Daniel no llegó a publicar un artículo académico sobre este particular.

La nutrición es un tema recurrente cuando se busca el equilibrio entre grupos. El escorbuto, una enfermedad debilitante causada por la falta de vitamina C, hacía estragos en la armada británica. James Lind, cirujano en el navío *HMS Salisbury*, realizó experimentos en 1742 sobre la cura del escorbuto. Lind escogió 12 marineros con escorbuto y les asignó dietas idénticas. Luego los distribuyó en seis parejas y trató cada una de ellas con un suplemento dietético diferente que se añadía a la ración diaria de alimento. Uno de los suplementos extra consistía en dos naranjas y un limón (Lind creía que una die-

ta ácida podría curar el escorbuto). Aunque Lind no recurriera a la asignación aleatoria, y aunque sus muestras fueran pequeñas para los usos actuales, demostró tener un estilo precursor al elegir a los 12 sujetos de su estudio «lo más parecidos entre sí como me fue posible». Quienes ingirieron cítricos (los primeros *limeys*[14] británicos) se curaron de manera rápida e incontestable, un hallazgo empírico que les cambió la vida y que surgió como resultado de los datos de Lind, aunque su teoría original estuviera equivocada.[15]

Pasaron casi 150 años desde Lind hasta el primer uso experimental documentado de la asignación aleatoria. Lo hizo Charles Peirce, un filósofo y científico estadounidense que experimentaba con la capacidad de los sujetos para detectar pequeñas diferencias de peso. En una publicación de 1885 poco fascinante, pero significativa desde el punto de vista metodológico, Peirce y su alumno Joseph Jastrow explicaron cómo hacían variar las condiciones experimentales extrayendo cartas de una baraja.[16]

El concepto riguroso de experimento aleatorio controlado no apareció hasta comienzos del siglo xx, en el trabajo del estadístico y genetista sir Ronald Aylmer Fisher, que se dedicaba a analizar datos de experimentos agrícolas. Los experimentos de asignación aleatoria aparecen en la obra de Fisher de 1925 *Statistical Methods for Research Workers* («Métodos estadísticos para los trabajadores de la investigación») y se describen en detalle en su obra maestra *The Design of Experiments* («El diseño de experimentos»), publicada en 1935.[17]

Fisher tuvo muchas ideas buenísimas, y algunas malas. Aparte de explicar la importancia de la asignación aleatoria, fue el inventor del método estadístico de máxima verosimilitud. Junto con el maestro de la econometría Sewall Wright (y con J.B.S. Haldane), inició el campo de la teoría genética de poblaciones. Pero también fue un comprome-

[14] La costumbre de consumir cítricos o su jugo se extendió rápidamente en la armada real y esto hizo que los marineros británicos recibieran en las colonias americanas el apelativo despectivo de *lime-juicers* («exprimidores de limas»), abreviado luego a *limeys*, designación que con el tiempo se extendió a todos los británicos en América. *(N. de la T.)*

[15] El experimento de Lind se describe en Duncan P. Thomas, «Sailors, Scurvy, and Science», *Journal of the Royal Society of Medicine*, vol. 3, 1885, páginas 75-83.

[16] Charles S. Peirce y Joseph Jastrow, «On Small Differences in Sensation», *Memoirs of the National Academy of Sciences*, vol. 3, 1885, páginas 75-83.

[17] Ronald A. Fischer, *Statistical Methods for Research Workers*, Oliver and Boyd, 1925, y Ronald A. Fisher, *The Design of Experiments*, Oliver and Boyd, 1935.

tido partidario de la eugenesia y defensor de la esterilización forzosa (al igual que el maestro de la regresión sir Francis Galton, quien acuñó el término *eugenesia*). Fisher, un gran fumador en pipa durante toda su vida, erró también en el debate acerca de la relación entre tabaco y salud, en parte debido a su arraigada creencia de que el gusto por fumar y el cáncer de pulmón compartían una causa genética común. Los efectos perjudiciales del tabaco sobre la salud parecen ahora bien fundamentados, aunque Fisher tuviera razón en su preocupación por los sesgos de selección en las investigaciones médicas. Se ha comprobado que muchas opciones de estilo de vida, como las dietas bajas en grasa o la ingestión de vitaminas, carecen de relación con mejoras en la salud cuando se evalúan por medio de experimentos aleatorios.

## Apéndice: Dominar la inferencia

Joven Caine: Estoy desconcertado.
Maestro Po: He ahí el comienzo de la sabiduría.

*Kung Fu*, segunda temporada, episodio 25

Este es el primero de una serie de apéndices dedicados a ampliar detalles econométricos y estadísticos esenciales. Se puede dedicar toda una vida al estudio de la inferencia estadística, y así lo hacen muchos maestros. Aquí ofrecemos un bosquejo breve de las ideas esenciales y las herramientas estadísticas básicas, suficiente para comprender tablas como las que aparecen en este capítulo.

El HIE (Health Insurance Experiment, «Experimento de Seguros Médicos») se basa en una muestra de participantes extraída (más o menos) al azar de entre la población elegible para el experimento. Si se extrajera una muestra distinta de la misma población se obtendrían resultados algo diferentes, pero la visión general seguiría siendo la misma siempre que la muestra fuera lo bastante grande como para que actuase la ley de los grandes números. ¿Cómo saber si unos resultados estadísticos constituyen un indicio firme, o si son poco más que el resultado de una baza afortunada con poca probabilidad de repetirse al tomar otra muestra? ¿Qué variabilidad cabe esperar entre las muestras? Las herramientas de

la inferencia estadística formal responden estas preguntas. Estas armas funcionan en todas las estrategias econométricas que nos interesan en esta obra. Cuantificar la incertidumbre asociada a una determinada muestra es un paso necesario en cualquier proyecto empírico y forma parte del camino hacia la comprensión de las afirmaciones de carácter estadístico emitidas por otros. Pasemos a explicar el concepto básico de la inferencia en el contexto de los efectos del tratamiento en el HIE.

La tarea que tenemos entre manos consiste en cuantificar la incertidumbre asociada a la media tomada sobre una muestra concreta y, en especial, a las medias de grupos y las diferencias entre ellas. Por ejemplo, nos gustaría saber si las grandes diferencias en gastos médicos entre los grupos de tratamiento del HIE pueden descartarse como producto del mero azar. Las muestras del HIE se extrajeron de un conjunto de datos mucho mayor que creemos que cubre la población de interés. La población del HIE consiste en todas las familias elegibles para el experimento (demasiado jóvenes para entrar en Medicare, etcétera). En lugar de estudiar los muchos millones de tales familias, se selecciona de entre ellas, al azar, un grupo mucho menor de unas 2.000 familias (y que contiene unas 4.000 personas), que luego se asignan al azar a cada uno de los 14 planes o grupos de tratamiento. Nótese que aquí intervienen dos tipos de asignación aleatoria: la primera en relación con la creación de la muestra del estudio, y la segunda con respecto al modo en que se asigna el tratamiento entre los integrantes de esa muestra. El *muestreo aleatorio* y la *asignación aleatoria* constituyen dos ideas muy relacionadas, pero diferentes.

### *Un mundo sin sesgos*

Cuantifiquemos primero la incertidumbre inducida por el muestreo aleatorio, empezando con una media simple de esa muestra como, por ejemplo, la salud promedio de todos los sujetos incluidos, medida a través de un índice de salud. Nuestro objetivo es reproducir la misma media para toda la población, es decir, el promedio evaluado sobre todos los individuos de la población de interés. Como ya se indicó en la página 14, la media de una variable sobre una población se

llama su *valor esperado* (o *esperanza matemática*). Representamos el valor esperado de una variable $Y_i$ como $E[Y_i]$. El valor esperado guarda una relación íntima con el concepto formal de *probabilidad*. El valor esperado puede escribirse como una media ponderada de todos los valores posibles que puede adoptar la variable $Y_i$, con los pesos consistentes en la probabilidad de que cada uno de estos valores aparezca en la población. En el ejemplo del lanzamiento de un dado, los pesos son iguales entre sí y valen todos 1/6 (véase el apartado 1.1).

A diferencia de lo que ocurre con nuestra notación para los valores medios, el símbolo del valor esperado no incluye ninguna referencia al tamaño de la muestra. Esto es así porque los valores esperados son cantidades que describen la población y que se definen sin hacer referencia a ninguna muestra concreta de individuos. Para una población dada existe un único valor esperado $E[Y_i]$ y, sin embargo, existen muchas medias $Avg_n[Y_i]$, dependiendo de cómo se elija $n$ y de qué individuos terminen por formar parte de la muestra. Como $E[Y_i]$ es un rasgo fijo de cada población concreta, nos referimos a esta cantidad como un *parámetro*. Las cantidades que varían de una muestra a otra, como la media sobre una muestra, se llaman *indicadores estadísticos muestrales* (o, simplemente, *indicadores estadísticos*, o *estadísticos muestrales*).

En este punto puede ser de ayuda pasar de la notación $Avg_n[Y_i]$ a otra más compacta para representar las medias, $\overline{Y}$. Obsérvese que prescindimos del subíndice $n$ para aligerar la escritura, de manera que a partir de ahora es responsabilidad suya el recordar que las medias muestrales se calculan para muestras de un tamaño determinado. La media muestral, $\overline{Y}$, constituye un buen estimador de $E[Y_i]$ (en estadística, un *estimador* es una función de los datos muestrales que se utiliza para estimar parámetros). El hecho de que esto sea así se justifica mediante la ley de los grandes números, que nos dice que, para muestras grandes, es muy probable que la media muestral sea muy similar a la media de la población correspondiente. Una propiedad relacionada con la anterior indica que el valor esperado de $\overline{Y}$ es también $E[Y_i]$. En otras palabras, si tomáramos infinitas muestras aleatorias, al extraer la media de todas las $\overline{Y}$ resultantes obtendríamos la media de la población subyacente. Cuando un indicador estadístico posee un valor esperado igual al parámetro correspondiente de la población, se dice que se trata de un *estimador insesgado* de ese pa-

rámetro. He aquí cómo se expresaría de manera formal el carácter insesgado de la media muestral:

Carácter insesgado de la media muestral: $E[\bar{Y}] = E[Y_i]$

No debe esperarse que la media muestral coincida exactamente con la correspondiente media de la población: la media de una determinada muestra puede quedar por encima, mientras que la de otra muestra puede quedar por debajo. El carácter insesgado indica que estas desviaciones no tenderán a caer de manera sistemática más hacia un lado que hacia otro, sino que, al tomar muchas muestras, al final se compensarán unas con otras. El carácter insesgado constituye una afirmación diferente a la ley de los grandes números, la cual dice que la media muestral se hace cada vez más parecida a la media de la población a medida que crece el tamaño de la muestra. El carácter insesgado de la media muestral es válido para muestras de cualquier tamaño.

### *Medida de la variabilidad*

No sólo nos interesan las medias, sino también la variabilidad de los datos. Para medir la variabilidad se suele tener en cuenta la desviación cuadrática media en torno al promedio, un proceso que asigna pesos iguales a las desviaciones positivas y a las negativas. La medida de la variabilidad resultante recibe el nombre de *varianza* o *desviación típica*.

La *varianza muestral de $Y_i$ en una muestra de tamaño n* se define como

$$S(Y_i)^2 = \frac{1}{n}\sum_{i=1}^{n} (Y_i - \bar{Y})^2.$$

La correspondiente varianza de la población se define considerando valores esperados en lugar de las medias, lo que da:

$$V(Y_i) = E\left[(Y_i - E[Y_i])^2\right]$$

Al igual que ocurre con $E[Y_i]$, la cantidad $V(Y_i)$ es un rasgo fijo de una población dada: un parámetro. Por eso es costumbre represen-

tarlo con una legra griega, $V(Y_i) = \sigma_Y^2$, que suele leerse como «sigma-i-cuadrado».[18]

Como las varianzas elevan los datos al cuadrado, pueden alcanzar valores muy grandes. Si se multiplica una variable por 10, entonces su varianza crece en un factor 100. Por eso a veces se describe la variabilidad por medio de la raíz cuadrada de la varianza: el resultado se denomina *desviación típica* (o *desviación estándar*), se escribe como $\sigma_Y$, y tiene como contrapartida muestral $S(Y_i)$, la raíz cuadrada de $S(Y_i)^2$.

La varianza constituye un hecho descriptivo acerca de la distribución de la variable $Y_i$. (Recuerde que la *distribución* de una variable consiste en el conjunto de valores que adopta esa variable, y la frecuencia relativa con que se observa cada valor en la población, o se genera por un proceso aleatorio.) Algunas variables sólo pueden adoptar un conjunto reducido de valores (como las variables binarias que se usan para marcar a las familias que cuentan con seguro médico), mientras que otras (como los ingresos) tienden a presentar un abanico amplio de posibilidades y mostrar algunos valores muy elevados mezclados con muchos otros más pequeños.

Es importante documentar la variabilidad de las variables con las que se trabaja. Pero nuestro objetivo ahora va algo más allá. Nos interesa cuantificar la varianza de las medias muestrales cuando se toman diversas muestras. Como el valor esperado de la media muestral es $E[Y_i]$ (por su carácter insesgado), la varianza de la población formada por las medias muestrales se puede escribir como

$$V(\bar{Y}) = E\left[(\bar{Y} - E[\bar{Y}])^2\right] = E\left[(\bar{Y} - E[Y_i])^2\right].$$

La varianza de un indicador estadístico como la media muestral es distinta de la varianza que se emplea con fines descriptivos. Escribimos $V(\bar{Y})$ para la varianza de la media muestral, pero $V(Y_i)$ (o bien $\sigma_Y^2$) para la varianza de los datos subyacentes. Como la cantidad $V(\bar{Y})$

[18] Las varianzas muestrales tienden a subestimar las varianzas de población. Por eso en ocasiones la varianza muestral se define como

$$S(Y_i)^2 = \frac{1}{i=1}\sum_{i=1}^{n} (Y_i - \bar{Y})^2,$$

es decir, dividiendo entre $n - 1$ en lugar de entre $n$. Esta fórmula modificada proporciona un estimador no sesgado de la varianza de la población correspondiente.

mide la variabilidad de un indicador estadístico calculado para múltiples muestras, en contraste con la dispersión de los datos de partida, $V(\bar{Y})$ recibe un nombre especial: varianza de muestreo.

La varianza de muestreo está relacionada con la varianza descriptiva pero, a diferencia de la varianza descriptiva, la varianza de muestreo viene determinada por el tamaño de la muestra. Veamos esto simplificando la fórmula de $V(\bar{Y})$. Primero sustituimos la expresión de $\bar{Y}$ dentro de la notación de la varianza:

$$V(\bar{Y}) = V\left(\left[\frac{1}{n}\sum Y_i\right]\right).$$

Para simplificar esta expresión tenemos en cuenta que el muestreo aleatorio garantiza que las observaciones individuales dentro de cada muestra no guardan relación entre ellas o, en otras palabras, que son estadísticamente independientes. Esta propiedad crucial nos permite aprovechar el hecho de que la varianza de una suma de observaciones estadísticamente independientes, cada una de ellas extraída al azar a partir de la misma población, es la suma de sus varianzas. Es más, dado que todos los $Y_i$ proceden de la misma población, cada elección tiene la misma varianza, $\sigma_Y^2$. Finalmente recurrimos a la propiedad de que la varianza de una constante (como $1/n$) multiplicada por $Y_i$ es igual al cuadrado de esa constante multiplicado por la varianza de $Y_i$. Teniendo en cuenta todo esto obtenemos:

$$V(\bar{Y}) = V\left(\left[\frac{1}{n}\sum_{i=1}^{n} Y_i\right]\right) = \frac{1}{n^2}\sum \sigma_Y^2.$$

Si simplificamos algo más tenemos:

$$V(\bar{Y}) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma_Y^2 = \frac{n\sigma_Y^2}{n^2} = \frac{\sigma_Y^2}{n}. \tag{1.5}$$

Hemos mostrado que la varianza de muestreo de la media muestral depende de la varianza de las observaciones subyacentes, $\sigma_Y^2$, y del tamaño de la muestra, $n$. Quizá usted ya haya deducido que disponer de más datos significa menos dispersión entre las medias muestrales de múltiples muestras. De hecho, cuando el tamaño de la muestra es muy grande casi no hay dispersión en absoluto, porque al crecer $n$ decrece el cociente $\sigma_Y^2/n$. He aquí la ley de los grandes

números en funcionamiento: a medida que $n$ se acerca a infinito, la media muestral se aproxima a la media de la población, y la varianza de muestreo desaparece.

En la práctica se suele trabajar con la desviación típica de las medias muestrales, más que con su varianza. La desviación típica de un indicador estadístico como la media muestral recibe el nombre de error típico (o error estándar). El error típico de la media muestral se puede escribir como

$$SE(\overline{Y}) = \frac{\sigma_Y}{\sqrt{n}}. \tag{1.6}$$

Cualquier estimación que conste en este libro lleva asociado su error típico. Esto incluye las medias muestrales (para las cuales se emplea la fórmula de error típico de la ecuación [1.6]), diferencias entre medias muestrales (que se abordan más adelante en este apéndice), coeficientes de regresión (tratados en el capítulo 2), y variables instrumentales u otras estimaciones más complejas. Las fórmulas de los errores típicos pueden llegar a ser bastante complicadas, pero la idea siempre es sencilla. Insistimos en la importancia de no confundir los errores típicos con las desviaciones típicas de las variables subyacentes, porque aunque ambas cantidades estén íntimamente relacionadas, miden cosas distintas.

Un último paso en el camino de los errores típicos: la mayoría de las cantidades exhibidas por las poblaciones, incluyendo la desviación típica que aparece en el denominador de (1.6), se desconocen y sólo pueden estimarse. En la práctica, por tanto, al cuantificar la varianza muestral de una media muestral se trabaja con un *error típico estimado*. Éste se obtiene reemplazando $\sigma_Y$ por $S(Y_i)$ en la fórmula de $SE(\overline{Y})$. En concreto, el error típico estimado de la media muestral se puede escribir como

$$\hat{SE}(\overline{Y}) = \frac{S(Y_i)}{\sqrt{n}}.$$

Solemos omitir el calificativo «estimado» al hablar de un estudio estadístico y sus errores típicos, pero es lo que tenemos en mente. Por ejemplo, los números que constan entre paréntesis en la tabla 1.4 son los errores típicos estimados para las diferencias de medias correspondientes.

### *La distribución* t *y el teorema del límite central*

Tras plantear de manera sencilla y esquemática el modo de medir la variabilidad por medio de los errores típicos, queda pendiente interpretar esta medida. La interpretación más simple recurre a la *distribución t*. Supongamos que los datos de que se dispone proceden de una distribución para la que creemos que la media de la población, $E[Y_i]$, adopta un valor particular, $\mu$ (léase esta letra griega como «mu»). Este valor constituye nuestra *hipótesis* de trabajo. El indicador estadístico $t$ para la media muestral bajo la hipótesis de trabajo de que $E[Y_i] = \mu$ se construye como

$$t(\mu) = \frac{\overline{Y} - \mu}{\hat{SE}\,(\overline{Y})}\,.$$

La hipótesis de trabajo es un punto de referencia que suele llamarse *hipótesis nula*. Cuando la hipótesis nula implica que $\mu = 0$, entonces el indicador $t$ se reduce al cociente entre la media muestral y su error típico estimado.

Muchas personas piensan que la ciencia de la inferencia estadística es aburrida, pero en realidad se acerca a lo milagroso. Un hecho estadístico milagroso es que si $E[Y_i]$ fuera realmente igual a $\mu$, entonces (siempre que la muestra sea lo bastante grande), la cantidad $t(\mu)$ presenta una distribución muestral muy parecida a la distribución normal gaussiana con forma de campana que se representa en la figura 1.1. Esta propiedad, que rige con independencia de que la propia distribución de $Y_i$ sea o no normal, recibe el nombre de *teorema del límite central* (TLC). El TLC permite adoptar decisiones sobre bases empíricas, acerca de si los datos disponibles apoyan, o ponen en duda, la hipótesis de que $E[Y_i]$ es igual a $\mu$.

El TLC arroja un resultado poderoso e impactante. Entre otras cosas implica que la distribución del indicador $t$ (para una muestra grande) es independiente de la distribución de los datos subyacentes empleados para calcularlo. Por ejemplo, supongamos que medimos el estado de salud empleando una variable binaria que diferencie las personas sanas de las enfermas, y que un 20% de la población esté enferma. La distribución de esta variable binaria tiene dos picos, uno de altura 0,8 en el valor 1 y otro de altura 0,2 en el valor 0. El TLC

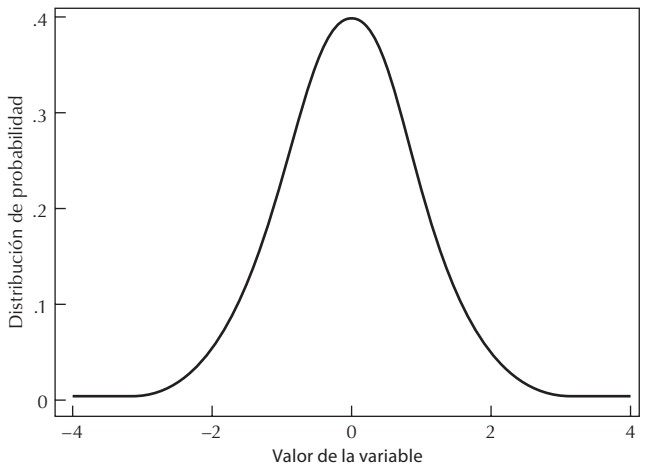



**Figura 1.1.** Una distribución normal gaussiana.

nos dice que, si se dispone de datos suficientes, la distribución *t* resultante tendrá una curvatura suave y con forma de campana, a pesar de que la distribución de los datos de partida exhiba tan sólo dos valores.

Veamos el TLC en acción en un experimento de muestreo. En los experimentos de muestreo se emplea un generador de números aleatorios por computadora para extraer muestras al azar de diferentes tamaños, una y otra vez. Hemos hecho esto para una variable binaria que adopta el valor uno el 80% de las veces, y para muestras con tamaños 10, 40 y 100. Calculamos el indicador *t* para cada tamaño de muestra y en un total de medio millón de muestras aleatorias, tomando siempre 0,8 como valor de $\mu$.

Las figuras 1.2 a 1.4 reflejan la representación gráfica de los 500.000 indicadores *t* calculados para cada uno de los tres tamaños de muestra del experimento, con una distribución normal superpuesta. Con sólo 10 observaciones resulta una distribución formada por picos, aunque se reconoce el perfil de una curva con forma de campana. A medida que se incrementa el tamaño de la muestra, mejora también el parecido con una distribución normal. El ajuste con la curva normal se vuelve casi perfecto con 100 observaciones.

La distribución normal gaussiana tiene media 0 y desviación típica 1. Los resultados con valor absoluto mayor que 2 se vuelven muy

improbables para cualquier variable normal. De hecho, tales realizaciones aparecen tan sólo un 5% de las veces. Como la distribución $t$ se parece tanto a una gaussiana normal, también esperamos que adopte valores entre +2 y –2 la mayor parte del tiempo. Por eso cualquier indicador $t$ con valor absoluto superior a 2 se suele considerar demasiado improbable como para ser compatible con la hipótesis nula que se haya usado para construirlo. Cuando la hipótesis nula es $\mu = 0$ y el valor absoluto de $t$ es mayor que 2, entonces se dice que la media muestral difiere significativamente de cero, o lo contrario si es menor que 2. Se emplean expresiones semejantes para otros valores de $\mu$.

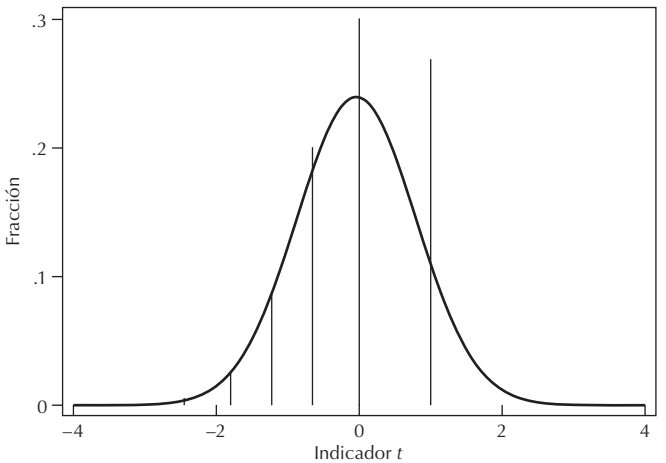



**Figura 1.2.** La distribución $t$ para la media de una muestra de tamaño 10.
*Nota:* Esta figura ilustra la distribución de la media muestral de una variable binaria que adopta el valor 1 con una probabilidad 0,8.

Podríamos considerar también la cuestión de la significancia estadística: en lugar de comprobar si la muestra es compatible con un valor específico de $\mu$, podríamos construir todo el conjunto de valores de $\mu$ que sean consistentes con los datos. El conjunto de tales valores se denomina intervalo de confianza para $E[Y_i]$. Cuando se calcula sobre repetidas muestras, el intervalo

$$\left[\overline{Y} - 2 \times \hat{SE}(\overline{Y}),\ \overline{Y} + 2 \times \hat{SE}\,(\overline{Y})\right]$$

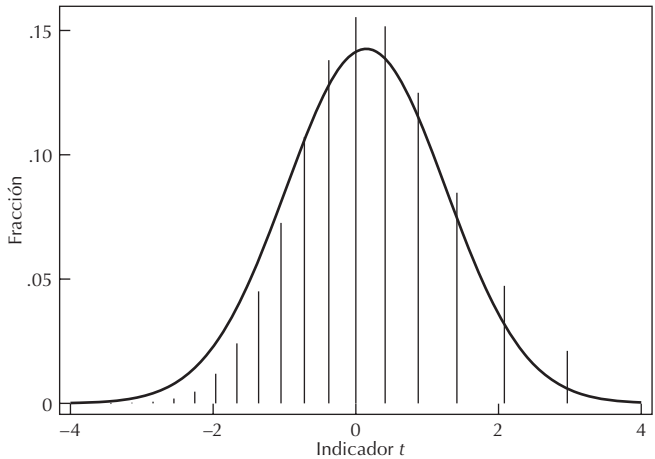



**Figura 1.3.** La distribución *t* para la media de una muestra de tamaño 40.
*Nota:* Esta figura ilustra la distribución de la media muestral de una variable binaria que adopta el valor 1 con una probabilidad 0,8.

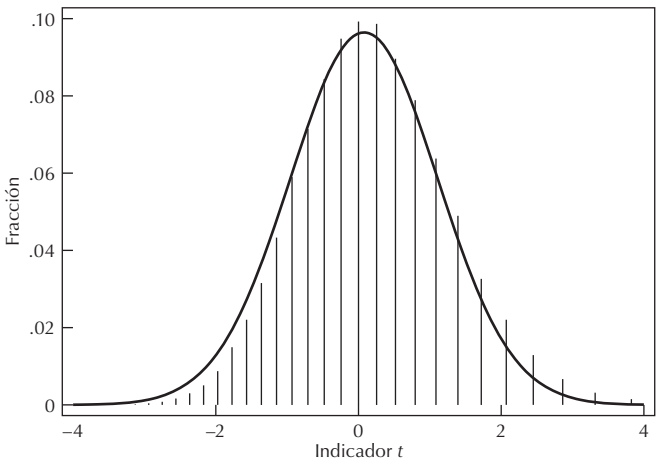



**Figura 1.4.** La distribución *t* para la media de una muestra de tamaño 100.
*Nota:* Esta figura ilustra la distribución de la media muestral de una variable binaria que adopta el valor 1 con una probabilidad 0,8.

debería contener $E[Y_i]$ en torno al 95% de las veces. Este intervalo se llama, por tanto, intervalo de confianza del 95% para la media de la población. Al describir el conjunto de valores de los parámetros consistentes con nuestros datos, los intervalos de confianza proporcionan un resumen compacto de la información que estos datos contienen acerca de la población de la que han sido extraídos.

### *Emparejamiento*[19]

Una media muestral es el número más solitario que quepa imaginar. Pero por suerte solemos manejar dos. Nos gusta muy en especial comparar las medias de los sujetos en los grupos de tratamiento y de control. Para referirnos a estas medias empleamos una notación compacta, $\overline{Y}^1$ para $Avg_n[Y_i|D_i = 1]$ e $\overline{Y}^0$ para $Avg_n[Y_i|D_i = 0]$. La media del grupo de tratamiento, $\overline{Y}^1$, es la media de las $n_1$ observaciones que pertenecen al grupo de tratamiento, e $\overline{Y}^0$ se define de un modo similar. El tamaño total de la muestra es $n = n_0 + n_1$.

Para nuestros fines, la diferencia entre $\overline{Y}^1$ e $\overline{Y}^0$ puede constituir, o bien una estimación del efecto causal del tratamiento (si $Y_i$ es un resultado), o una comprobación del equilibrio (si $Y_i$ es una covariable). Para centrar el foco de la explicación supondremos lo primero. La hipótesis nula más importante en este contexto es que el tratamiento no surte ningún efecto, en cuyo caso las dos muestras empleadas para construir las medias de tratamiento y de control provendrán de la misma población. Por otra parte, si el tratamiento altera los resultados, entonces las poblaciones de las que se extraen las observaciones de tratamiento y de control serán necesariamente distintas. En particular exhibirán distintas medias, que representaremos como $\mu^1$ y $\mu^0$.

Para decidir si los indicios favorecen la hipótesis de que $\mu^1 = \mu^0$ buscamos diferencias estadísticamente significativas entre las medias muestrales correspondientes. Los resultados estadísticamente significativos aportan indicios sólidos en favor de la efectividad del trata-

[19] En inglés se usa el término *matching* para este proceso, que en lenguas latinas suele traducirse como *emparejamiento*, aunque en general no consista en formar parejas, sino grupos más amplios (*muestras concordes* o *apareadas*, *matched samples*). *(N. de la T.)*.

miento, mientras que los resultados que no alcanzan tal significancia son compatibles con la idea de que la diferencia observada entre las medias de tratamiento y de control se debe al azar. En este contexto, «al azar» quiere decir que en un experimento hipotético que implicara muestras muy grandes (tan grandes que eliminaran de forma efectiva cualquier varianza muestral) encontraríamos que las medias de tratamiento y de control coinciden.

La significancia estadística se determina por medio del indicador *t* adecuado. Un ingrediente fundamental en cualquier receta de un indicador *t* es el error típico que reside en el denominador del cociente. El error típico de una comparación de medias es la raíz cuadrada de la varianza de muestreo de $\overline{Y}^1 - \overline{Y}^0$. Si se recurre al hecho de que la varianza de la diferencia de dos variables estadísticas independientes es la suma de sus varianzas, tenemos que

$$\begin{aligned} V(\overline{Y}^1 - \overline{Y}^0) &= V(\overline{Y}^1) + V(\overline{Y}^0) \\ &= \frac{\sigma_Y^2}{n_1} + \frac{\sigma_Y^2}{n_0} = \sigma_Y^2 \left[\frac{1}{n_1} + \frac{1}{n_0}\right]. \end{aligned}$$

La segunda de estas igualdades usa la ecuación (1.5), que da la varianza de muestreo de una media simple. El error típico que necesitamos es, por tanto,

$$SE(\overline{Y}^1 - \overline{Y}^0) = \sigma_Y \sqrt{\frac{1}{n_1} + \frac{1}{n_0}}.$$

Para deducir esta expresión se da por supuesto que las varianzas de las observaciones individuales son las mismas en los grupos de tratamiento y de control. Este supuesto permite emplear el símbolo $\sigma_Y^2$ para la varianza común. Una fórmula algo más complicada permitiría tener en cuenta varianzas distintas para cada grupo, incluso aunque las medias fueran las mismas (una idea que recuperaremos al tratar los errores típicos de la regresión robusta en el apéndice del capítulo 2).[20]

[20] Cuando se usan varianzas distintas para las observaciones de tratamiento y de control se tiene que

$$SE(\overline{Y}^1 - \overline{Y}^0) = \sqrt{\frac{V^1(Y_i)}{n_1} + \frac{V^0(Y_i)}{n_0}}.$$

donde $V^1(Y_i)$ es la varianza de las observaciones de tratamiento, y $V^0(Y_i)$ es la varianza de las observaciones de control.

Teniendo en cuenta que $\sigma_Y^2$ es una cantidad que sólo cabe estimar, en la práctica trabajamos con el error típico estimado:

$$\hat{SE}\,(\bar{Y}^1 - \bar{Y}^0) = S(Y_i)\sqrt{\frac{1}{n_1} + \frac{1}{n_0}}, \tag{1.7}$$

donde $S(Y_i)$ es la desviación típica muestral combinada. Se trata de la desviación típica muestral calculada usando los datos combinados de los dos grupos, tratamiento y control.

Bajo la hipótesis nula de que $\mu^1 - \mu^0$ es igual al valor $\mu$, el estimador $t$ para la diferencia de medias es:

$$t(\mu) = \frac{\bar{Y}^1 - \bar{Y}^0 - \mu}{\hat{SE}(\bar{Y}^1 - \bar{Y}^0)}.$$

Usamos este indicador $t$ para poner a prueba la hipótesis de trabajo acerca de $\mu^1 - \mu^0$, y para construir los intervalos de confianza para esta diferencia. Cuando la hipótesis nula consiste en que las medias son iguales ($\mu = 0$), el indicador $t(\mu)$ resulta igual a la diferencia de las medias muestrales dividido por el error típico estimado de esta diferencia. Cuando el indicador $t$ es lo bastante grande como para descartar que la diferencia sea nula, decimos que la diferencia estimada es estadísticamente significativa. El intervalo de confianza de una diferencia de medias es la diferencia de las medias muestrales más o menos dos veces los errores típicos.

Tengamos presente que los indicadores $t$ y los intervalos de confianza dicen poco acerca de si los resultados observados son de por sí grandes o pequeños. Se da un indicador $t$ elevado cuando el efecto estimado de interés es grande, pero también cuando el error típico correspondiente es pequeño (como ocurre cuando se tiene la suerte de contar con una muestra amplia). Del mismo modo, la amplitud de un intervalo de confianza queda determinada por la precisión estadística que se refleja en los errores típicos, y no por la magnitud de las relaciones que intentamos desvelar. Y, al revés, un indicador estadístico $t$ puede resultar bajo porque la diferencia de medias estimadas también lo sea, o porque el error típico de esta diferencia sea grande. El hecho de que una diferencia estimada no sea significativamente diferente de cero no tiene por qué implicar que la relación que se está

investigando sea pequeña o irrelevante. La falta de significancia estadística suele reflejar una falta de precisión estadística, es decir, una gran varianza muestral. Los maestros tienen en cuenta este hecho a la hora de discutir resultados econométricos.

2

# Regresión

Kwai Chang Caine: Al trabajador se lo conoce por sus herramientas. La pala para el cavador. El hacha para el leñador. El econometrista usa regresiones.

*Kung Fu*, primera temporada, episodio 8

## *Nuestro camino*

Cuando encontramos cortado el camino de la asignación aleatoria, buscamos rutas alternativas hacia el conocimiento causal. Si se esgrimen con destreza, las armas econométricas distintas a la asignación aleatoria pueden tener una capacidad para revelar causas muy similar a la de un experimento real. La más básica de estas técnicas es la regresión, que compara sujetos de tratamiento y de control con las mismas características observadas. Los conceptos de la regresión son de carácter fundamental, y sirven de base a herramientas más elaboradas que se describen en capítulos posteriores. La inferencia causal basada en la regresión parte del supuesto de que, una vez se han igualado las variables observadas en los grupos de tratamiento y de control, entonces también queda eliminado en su mayor parte el sesgo de selección debido a cosas que no podemos ver. Ilustramos esta idea con una investigación empírica del beneficio económico que reporta la formación en centros privados de élite.

## 2.1 Historia de dos universidades

Los estudiantes que cursaban carreras de cuatro años en centros privados de Estados Unidos pagaron en promedio unos 29.000 dólares en concepto de matrícula y tasas en el año académico 2012-2013. Quienes acudieron a una universidad pública en su estado natal pagaron menos de 9.000. Una educación privada de élite podría ser mejor por varias razones: grupos más reducidos en cada clase, instalaciones deportivas más nuevas, profesorado más distinguido y estudiantes más listos. Pero 20.000 dólares por año académico suponen una gran diferencia. Cabría preguntarse si vale la pena.

La cuestión de juntar manzanas con manzanas correspondería, en este caso, a preguntarse cuáles serían los ingresos de un graduado de cuarenta años nacido en Massachusetts y graduado en Harvard si hubiera cursado sus estudios en la Universidad de Massachusetts (U-Mass). El dinero no lo es todo pero, como decía Groucho Marx: «El dinero te libra de hacer las cosas que no te gustan. Como a mí no me gusta hacer casi nada, el dinero me viene muy bien». Así que cuando nos preguntamos si vale la pena el gasto adicional que supone un centro privado, nos centramos en la posible mejora de ingresos que quizá disfruten quienes estudiaron en universidades privadas de élite. Puede haber otras razones, y no sólo un aumento de ingresos, para preferir una institución privada de élite en lugar de la universidad local del estado. Muchos estudiantes universitarios conocen a sus futuros cónyuges y forjan amistades para toda la vida en la facultad. Aun así, cuando una familia invierte 100.000 dólares adicionales, o más, en la formación del capital humano, parece verosímil que las expectativas de unos mayores ingresos en el futuro formen parte del asunto.

La comparación de ingresos entre quienes asistieron a universidades de distintos tipos siempre revela grandes diferencias en favor del alumnado de centros de élite. Aunque, pensándolo bien, es fácil ver por qué es poco probable que la comparación de los ingresos de quienes estudiaron en Harvard con los de quienes lo hicieron en U-Mass revele las ventajas de tener un título de Harvard. La comparación refleja que los titulados de Harvard suelen tener calificaciones más altas en secundaria y notas mejores en las pruebas de admisión, suelen estar más motivados y quizá cuenten con otras habilidades y talentos. No pretendemos ofender a los muchísimos buenos estudiantes que

acuden a U-Mass, pero es endiabladamente difícil entrar en Harvard, y quienes lo logran conforman un grupo especial y selecto. En contraste, U-Mass acepta, e incluso subvenciona, a casi cualquier solicitante de Massachusetts que haya sacado unas notas decentes en la enseñanza media. Por lo tanto, cabría esperar que la comparación de ingresos entre titulados de estas universidades estuviera contaminada por un sesgo de selección, como las comparaciones entre tipos de seguro médico que tratamos en el capítulo anterior. También hemos visto que este tipo de sesgos de selección se elimina con la asignación aleatoria. Pero, por desgracia, las oficinas de Harvard aún no están preparadas para cambiar sus criterios de admisión por un generador de números aleatorios.

La trascendencia de la elección de universidad deberá esclarecerse recurriendo a los datos que generan las decisiones habituales en el proceso de solicitud, admisión y matriculación tomadas por estudiantes y universidades de varios tipos. ¿Cabe emplear estos datos para simular el experimento aleatorio que nos habría gustado ejecutar en este contexto? No de manera perfecta, sin duda, pero quizá podamos acercarnos. La clave de este desafío radica en que muchas decisiones, incluidas las que tienen que ver con la elección de facultad, incorporan cierta cantidad de variación debida al azar causada por consideraciones económicas, circunstancias personales y tiempo.

El azar se puede explotar si se identifica una muestra de solicitantes situados al filo de la navaja, que fácilmente podrían haber caído tanto a un lado como al otro. ¿Hay algún caso de estudiante admitido en Harvard pero que al final terminara acudiendo a su universidad estatal local? A nuestra amiga y antigua estudiante de doctorado del MIT Nancy, le ocurrió justo eso. Nancy se crió en Tejas, así que su universidad estatal era la Universidad de Tejas (UT). El buque insignia de la UT, el campus de Austin, aparece catalogado como «muy competitivo» en la clasificación de Barron's, pero no es Harvard. Sin embargo, la UT es mucho más barata que Harvard (la revista *The Princeton Review* calificó el campus de la UT en Austin como el mejor en cuanto a relación calidad-precio). A Nancy la admitieron tanto en Harvard como en la UT, pero acabó eligiendo la UT porque su oficina de admisiones, ansiosa por mejorar la media de las notas de ingreso en el campus, ofreció a Nancy y a otros cuantos solicitantes destacados un paquete de ayudas económicas especialmente generoso, lo que Nancy aceptó encantada.

¿Qué consecuencias tuvo para Nancy la decisión de aceptar la oferta de la UT y rechazar la de Harvard? Las cosas le han ido bastante bien a pesar de haber elegido la UT en lugar de Harvard: hoy es profesora de economía en otra universidad de la Ivy League[1] en Nueva Inglaterra. Pero esto es una única observación. Bueno, en realidad tenemos dos observaciones, porque nuestra amiga Mandy se graduó en la Universidad de Virginia, su estado de origen, tras rechazar ofertas de Duke, Harvard, Princeton y Stanford. Ahora Mandy es profesora en Harvard.

Un tamaño de muestra igual a dos es demasiado poco para extraer una inferencia causal fiable. Nos gustaría comparar a muchas personas como Mandy y Nancy con otras parecidas, pero que eligieron las universidades privadas. Cabe la esperanza de obtener conclusiones de validez general a partir de comparaciones entre grupos mayores. Pero el acceso a una muestra grande no basta. El paso primero y más importante en nuestro esfuerzo por aislar la componente que es fruto del azar en la elección de universidad consistiría en mantener constantes las diferencias más evidentes e importantes entre los estudiantes que acuden a universidades públicas y los que acuden a las privadas. De este modo aspiramos (aunque no lo podemos prometer) a que *el resto permanezca igual*.

Veamos un ejemplo numérico que recurre a una muestra pequeña para ilustrar el concepto de *ceteris paribus* (usaremos más datos cuando llegue el momento de hacer el trabajo empírico real). Supongamos que las dos únicas circunstancias relevantes en la vida, al menos en lo que respecta al nivel de ingresos, fueran las notas en los exámenes de acceso a la universidad y en qué universidad se cursan los estudios. Pensemos en Uma y Harvey, ambos con la misma calificación combinada en matemáticas y lectura de 1400 en las pruebas de acceso a la universidad.[2] Uma fue a la U-Mass, mientras que Harvey acudió a Harvard. Comparemos en primer lugar los ingresos de Uma y Harvey. Como hemos supuesto que lo único que importa al

[1] Ivy League es el nombre informal con que se conoce a un conjunto de ocho universidades privadas muy prestigiosas del nordeste de EE. UU. (Brown, Columbia, Cornell, Dartmouth College, Harvard, Pensilvania, Yale y Princeton). *(N. de la T.)*

[2] Aquí se sigue el esquema de puntuación de las pruebas de acceso a la universidad de Estados Unidos (SAT) anterior al año 2005, según el cual en las notas finales se suman los puntos obtenidos en matemáticas y en expresión oral, cada una de las cuales toma valores entre 0 y 800, de modo que el máximo combinado asciende a 1600.

respecto es, aparte de la universidad elegida, la puntuación final en las pruebas de acceso, la comparación de Uma con Harvey se produce en condiciones *ceteris paribus*.

En la práctica, por supuesto, la vida es más complicada. Este ejemplo sencillo ya plantea una dificultad notable: Uma es una mujer, mientras que Harvey es un hombre. Las mujeres con un mismo nivel de estudios que los hombres suelen ganar menos dinero que éstos, quizá por discriminación o debido al tiempo que pasan fuera del mercado laboral para tener hijos. El hecho de que Harvey gane un 20% más que Uma podría deberse al efecto de una formación mejor en Harvard, pero también podría reflejar una diferencia entre hombres y mujeres debida a otros motivos.

Nos gustaría desenredar el efecto Harvard, puro, de entre todos esos otros motivos. Sería fácil si lo único relevante fuera el género: se sustituye a Harvey por una estudiante femenina de Harvard, Hanna, que también haya sacado 1400 en las pruebas de acceso a la universidad, y se la compara con Uma. Al final, y dado que perseguimos conclusiones generales que van más allá de las historias individuales, buscamos muchas más parejas formadas por estudiantes de las dos universidades y que cumplan el criterio de ser iguales en cuanto a género y nota de acceso. Es decir, calculamos la media de la diferencia de ingresos entre estudiantes de Harvard y de U-Mass con géneros y notas de acceso iguales. La media de estas diferencias específicas del grupo Harvard frente a U-Mass constituyen nuestra primera acción para intentar estimar el efecto causal de la formación en Harvard. Este es un estimador econométrico *apareado* que *controla* (es decir, mantiene fijos) el género y la nota de acceso. Si se acepta que todos los estudiantes que acuden a Harvard y a U-Mass tienen el mismo potencial de ingresos, condicionado por el género y la nota de acceso, entonces este estimador capta el efecto causal promedio que ejerce sobre los ingresos el hecho de graduarse en Harvard.

### *El emparejador que las empareje*

Pero resulta que hay muchas otras cosas que influyen en los ingresos, aparte del género, la universidad o la puntuación obtenida en las pruebas de acceso. Como las decisiones sobre a qué universidad

acudir no se adoptan de manera aleatoria, hay que controlar todos los factores que determinan tanto la elección de universidad como los ingresos posteriores. Entre estos factores se cuentan ciertas características del estudiante, como la destreza para expresarse por escrito, la diligencia, las relaciones familiares, y más. Intentar el control de un abanico tan amplio de factores parece una tarea desalentadora: las posibilidades son infinitas, y muchas de las características resultan difíciles de cuantificar. Pero Stacy Berg Dale y Alan Krueger dieron con un atajo ingenioso y fascinante.[3] En lugar de identificar todo lo que podría influir en la elección de universidad y en los ingresos, trabajan con una medida conjunta clave: las características de las universidades en las que los estudiantes presentaron las solicitudes y fueron admitidos.

Consideremos de nuevo la historia de Uma y Harvey: ambos presentaron solicitud y fueron admitidos tanto en U-Mass como en Harvard. El hecho de que Uma pidiera Harvard indica que tenía alguna motivación para ir allí, mientras que su admisión significa que tenía las capacidades necesarias para triunfar allí, como Harvey. Al menos eso es lo que piensa la oficina de admisiones de Harvard, y no es fácil engañarlos.[4] Sin embargo, Uma opta por formarse en U-Mass, por ser más barata. Su elección podría deberse a factores que no están relacionados con su potencial de ingresos, como tener un tío muy exitoso que hubiera estudiado en U-Mass, que uno de sus mejores amigos eligiera U-Mass, o a que se le pasara el plazo para solicitar una beca del Rotary Club que habría conseguido con facilidad y que le habría costeado los estudios en una universidad de la Ivy League, más cara. Si sucesos azarosos de este estilo fueron decisivos tanto para Uma como para Harvey, entonces los dos forman una buena pareja.

---

[3] Stacy Berg Dale y Alan B. Krueger, «Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables», *Quarterly Journal of Economics*, vol. 117, número 4, noviembre de 2002, páginas 1491-1527.

[4] Lo que no quiere decir que sea imposible. Adam Wheeler logró acceder a Harvard de manera ilícita con expedientes y calificaciones manipulados en 2007. A pesar del engaño, las calificaciones que obtenía Adam en Harvard eran sobre todo notables y sobresalientes hasta que se descubrió la trampa (John R. Ellement y Tracy Jan, «ExHarvard Student Accused of Living a Lie», *The Boston Globe*, 18 de mayo de 2010).

Dale y Krueger analizaron un gran conjunto de datos denominado *Universidad y Más Allá* (*College and Beyond*, C&B). Los datos C&B contienen información sobre miles de estudiantes que ingresaron en un conjunto de universidades de Estados Unidos con una política de selección entre moderada y alta, junto con información contextual aportada por los propios estudiantes al pasar las pruebas de acceso (lo que sucede alrededor de un año antes de entrar en la universidad) e información recogida en 1996 (mucho después de que la mayoría hubiera obtenido ya sus títulos académicos). Nuestro análisis se centra en los estudiantes que ingresaron en 1976 y que estaban trabajando en 1995 (la mayoría de los graduados universitarios adultos tiene trabajo). Entre los centros de estudios se cuentan universidades privadas prestigiosas, como la Universidad de Pensilvania, Princeton o Yale; un conjunto de universidades privadas más pequeñas, como Swarthmore, Williams u Oberlin, y cuatro universidades públicas (Michigan, la Universidad de Carolina del Norte, Penn State y la Universidad Miami en Ohio). La puntuación media (1978) en las pruebas de acceso a la universidad en estos centros va desde un mínimo de 1020 en Tulane hasta un máximo de 1370 en Bryn Mawr. Los costes de matriculación en 1976 ascendían a tan sólo 540 dólares en la Universidad de Carolina del Norte, y la considerable cifra de 3.850 dólares en Tufts (así estaban las cosas por entonces).

La tabla 2.1 presenta una versión simplificada de la estrategia de emparejamiento de Dale y Krueger, expuesta en lo que llamamos «matriz universitaria de emparejamiento». La tabla consigna decisiones de solicitud, admisión y matriculación para una lista (ficticia) de nueve estudiantes, cada uno de los cuales solicitó el ingreso en hasta tres centros elegidos de una lista imaginaria de seis. Tres de estos centros son públicos (Omniópolis, Altópolis y Otrópolis) y tres privados (Treposa, Foliosa y Astutosa). Cinco de nuestros nueve estudiantes (números 1, 2, 4, 6 y 7) acudieron a centros privados. Los ingresos anuales medios de este grupo ascienden a 92.000 dólares. Los otros cuatro, con ingresos medios de 72.500, estudiaron en centros públicos. La diferencia de casi 20.000 dólares entre estos dos grupos parece indicar que acudir a una universidad privada proporciona una ventaja considerable.

**Tabla 2.1. Matriz universitaria de emparejamiento**

| | | Centros privados | | | Centros públicos | | | |
|---|---|---|---|---|---|---|---|---|
| Grupo de solicitantes | Estudiante | Treposa | Foliosa | Astutosa | Omniópolis | Altópolis | Alterópolis | Ingresos en 1996 |
| A | 1 | | Rechazo | Admisión | | Admisión | | 110,000 |
| | 2 | | Rechazo | Admisión | | Admisión | | 100,000 |
| | 3 | | Rechazo | Admisión | | Admisión | | 110,000 |
| B | 4 | Admisión | | | Admisión | | Admisión | 60,000 |
| | 5 | Admisión | | | Admisión | | Admisión | 30,000 |
| C | 6 | | Admisión | | | | | 115,000 |
| | 7 | | Admisión | | | | | 75,000 |
| D | 8 | Rechazo | | | Admisión | Admisión | | 90,000 |
| | 9 | Admisión | | | Admisión | Admisión | | 60,000 |

*Nota:* Las decisiones de ingreso se resaltan en gris.

Los estudiantes de la tabla 2.1 se distribuyen en cuatro grupos definidos por el conjunto de centros que solicitaron y en los que resultaron admitidos. Cabe esperar que los estudiantes incluidos en cada grupo tengan ambiciones similares en cuanto a carrera profesional, y que hayan sido considerados de capacidades semejantes por el personal encargado de la admisión en los centros solicitados. De este modo, las comparaciones dentro de los grupos deberían acercarse más a juntar manzanas con manzanas que las comparaciones no controladas que mezclan a todos los estudiantes.

Los estudiantes del grupo A pidieron el ingreso en dos centros privados, Foliosa y Astutosa, y en uno público, Altópolis. Aunque estos estudiantes fueran rechazados en Foliosa, lograron la admisión tanto en Astutosa como en Altópolis. Los estudiantes 1 y 2 acudieron a Astutosa, mientras que el 3 eligió Altópolis. Los estudiantes del grupo A tienen ingresos elevados y es probable que procedan de familias de clase media alta (como lo indica el hecho de que pidieran el ingreso en más universidades privadas que públicas). El estudiante 3, aunque fue admitido en Astutosa, prefirió la opción de Altópolis, más barata, quizá para ahorrar dinero a su familia (como nuestras amigas Nancy y Mandy). Aunque a los estudiantes del grupo A les vaya bastante bien, con ingresos medios elevados y una proporción alta de estudios cursados en centros privados, el diferencial de la escuela privada es

negativo en este colectivo: (110 + 100)/2 – 110 = –5. En otras palabras, un salto de –5.000 dólares.

Este cotejo dentro del grupo A no es más que una de las muchas comparaciones que pueden establecerse a través de emparejamientos en la tabla. El grupo B incluye dos estudiantes, cada uno de los cuales solicitó el ingreso en una universidad privada y en dos públicas (Treposa, Omniópolis y Alterópolis). Los estudiantes del grupo B obtienen ingresos medios inferiores a los del A. Ambos lograron la admisión en las tres universidades que solicitaron. El número 4 ingresó en Treposa, mientras que el número 5 prefirió Alterópolis. El diferencial de ingresos aquí asciende a 30.000 dólares (60 – 30 = 30). Esta brecha sugiere una ventaja muy sustancial debida a la educación privada.

El grupo C incluye dos estudiantes que presentaron la solicitud en un solo centro (Foliosa), donde fueron admitidos y donde se matricularon. Los ingresos del grupo C no nos dicen nada acerca de las consecuencias de asistir a un centro privado, porque ambos estudiantes acudieron a una universidad privada. Los dos estudiantes del grupo D presentaron solicitudes en tres centros, fueron admitidos en dos y finalmente eligieron opciones distintas. Pero estos estudiantes se matricularon en Omniópolis y Altópolis, ambas universidades públicas, así que sus ingresos tampoco esclarecen el peso de la enseñanza privada. Los grupos C y D no aportan información porque, desde el punto de vista de nuestro afán por estimar el efecto del tratamiento «universidad privada», cada uno de ellos se compone, o bien de individuos todos ellos sujetos al tratamiento, o bien de individuos todos ellos del grupo de control.

En nuestro ejemplo el interés se centra en los grupos A y B, porque ambos incluyen sujetos que cursaron estudios en escuelas públicas y en escuelas privadas, y que solicitaron el ingreso y fueron admitidos en el mismo conjunto de centros. Para construir un único estimador que emplee todos los datos disponibles procedemos a promediar los estimadores específicos de cada grupo. El promedio de –5.000 (del grupo A) y 30.000 (del grupo B) es 12.500. He aquí una buena estimación del efecto que ejerce acudir a un centro privado sobre los ingresos medios porque, en gran medida, controla tanto las decisiones como las capacidades de los solicitantes.

La media simple de las diferencias tratamiento-control en los grupos A y B no es la única comparación bien controlada que puede

efectuarse con estos datos. Por ejemplo, cabe construir una media ponderada que refleje el hecho de que el grupo B incluye dos estudiantes, mientras que el grupo A tiene tres. En este caso la media ponderada se calcularía como

$$\left(\frac{3}{5} \times -5.000\right) + \left(\frac{2}{5} \times 30.000\right) = 9.000.$$

Al dar más peso a los grupos grandes, el sistema de ponderación empleado hace un uso más eficiente de los datos, lo que podría conducir a una síntesis estadística más precisa en lo que respecta al diferencial de ingresos público-privado.

El aspecto más importante en este contexto reside en la naturaleza de las comparaciones, que siguen el principio de colocar las manzanas con las manzanas, y las naranjas con las naranjas. Las manzanas del grupo A se comparan con otras manzanas del grupo A, mientras que las naranjas del grupo B se comparan sólo con naranjas. En contraste, las comparaciones simplonas que se limitan a poner a los alumnos de instituciones privadas frente a los de las públicas arrojan una diferencia mucho mayor, de 19.500 dólares, al incluir a todos los alumnos de la tabla. Incluso si se limita a los cinco estudiantes de los grupos A y B, una comparación no controlada conduce a un salto de 20.000 dólares (20 = (110 + 100 + 60)/3 – (110 + 30)/2). Estas comparaciones mayores y sin control reflejan un sesgo de selección: los estudiantes que solicitan los centros privados y que son admitidos en ellos tienen ingresos mayores con independencia de dónde decidan finalmente estudiar.

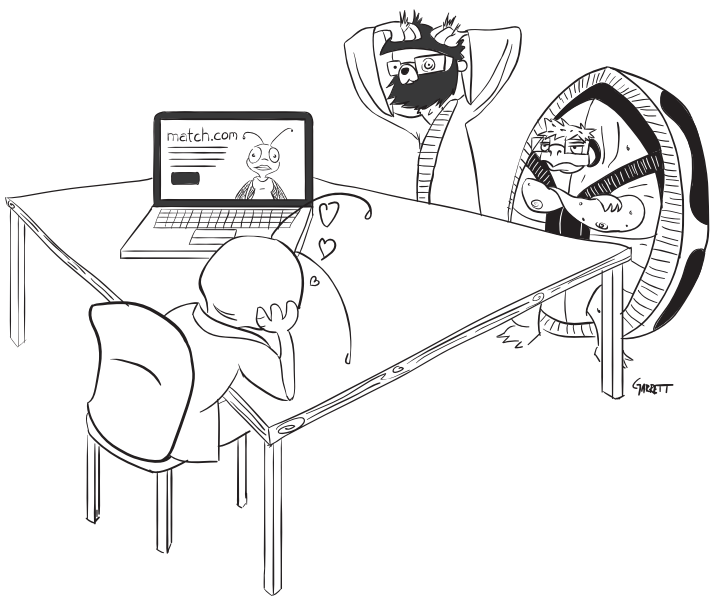
Los indicios de sesgos de selección surgen al comparar los ingresos medios entre los grupos A y B (y no dentro de ellos). Los ingresos medios del grupo A, donde dos tercios solicitaron centros privados, están en torno a 107.000 dólares. Los ingresos medios del grupo B, donde dos tercios solicitaron centros públicos, ascienden a tan sólo 45.000. Nuestras comparaciones dentro de cada grupo ponen de manifiesto que gran parte de esta diferencia carece de relación con el tipo de universidad a la que acudieron. Las diferencias entre grupos se explican más bien por una combinación de ambiciones y capacidades, como traslucen las decisiones sobre solicitudes y el conjunto de centros donde los alumnos fueron admitidos.

## 2.2 Emparéjame y hazme una regresión

La regresión es la herramienta preferida de los maestros, aunque sólo sea como punto de partida para aplicar, luego, estrategias empíricas más elaboradas. Aunque la regresión tiene esplendorosas virtudes, conviene contemplarla como una herramienta de emparejamiento automático. En concreto, la regresión brinda estimaciones que son promedios ponderados de múltiples comparaciones similares a las que construimos para los grupos de nuestra elemental matriz de emparejamientos (el apéndice de este capítulo trata una conexión estrecha entre la regresión y el valor esperado).

Los ingredientes fundamentales de la receta para la regresión son:

- la *variable dependiente*, en este caso los ingresos del estudiante $i$ en su vida posterior, también llamada la variable resultado (denotada como $Y_i$);
- la *variable de tratamiento*, en este caso una variable binaria que señala si los estudiantes acudieron a una universidad privada o a una pública (denotada como $P_i$), y
- un conjunto de *variables de control*, en este caso variables que identifican conjuntos de centros en los que los estudiantes presentan sus solicitudes y son admitidos o no.



En nuestra matriz de emparejamientos los cinco estudiantes de los grupos A y B (tabla 2.1) aportan datos útiles, mientras que los estudiantes de los grupos C y D se pueden descartar. En un conjunto de datos que contenga los que permanecen tras el descarte de los grupos C y D, basta una sola variable que señale a los estudiantes del grupo A para indicarnos a cuál de los grupos restantes pertenecen los estudiantes, porque los que no estén en A pertenecerán necesariamente a B. Esta variable, a la que llamaremos $A_i$, será nuestro único control. Obsérvese que tanto $P_i$ como $A_i$ son variables binarias (*dummy*), es decir, cuando valen 1 indican que las observaciones pertenecen a un cierto grupo o condición, y lo contrario cuando valen 0. Las variables binarias clasifican los datos en categorías simples del tipo sí/no. Aun así, si se crean muchas variables de este tipo se puede conseguir un conjunto de variables de control tan detallado como se desee.[5]

En este contexto un modelo de regresión es una ecuación que relaciona la variable de tratamiento con la variable dependiente, manteniendo las variables de control fijas al incluirlas en el modelo. Con sólo una variable de control, $A_i$, la regresión de interés podría escribirse como

$$Y_i = \alpha + \beta P_i + \gamma A_i + e_i. \tag{2.1}$$

La diferencia entre la variable de tratamiento, $P_i$, y la variable de control, $A_i$, en la ecuación (2.1) es conceptual, no formal: no hay nada en la ecuación (2.1) que indique cuál es cuál. Es la pregunta objeto de investigación y la estrategia empírica las que justifican la clasificación de las variables y determinan qué funciones desempeñan.

Como en el capítulo anterior, también aquí designamos con letras griegas los parámetros, para así distinguirlos de las variables del modelo. Los parámetros de la regresión (llamados *coeficientes de regresión*) son

- la ordenada en el origen, $\alpha$ (alfa);
- el efecto causal del tratamiento, $\beta$ (beta), y
- el efecto de pertenecer al grupo A, $\gamma$ (gamma).

[5] Cuando los datos pertenecen a uno de un total de J grupos, entonces se requieren J-1 variables binarias para una descripción completa de esos grupos. La categoría a la que no le corresponde ninguna variable binaria recibe el nombre de *grupo de referencia*.

El último término de la ecuación (2.1) es el *residuo*, $e_i$ (también llamado *término de error*). Los residuos se definen como la diferencia entre los $Y_i$ observados y los valores estimados que genera el modelo de regresión concreto que se emplee. Estos valores estimados se escriben como

$$\hat{Y}_i = \alpha + \beta P_i + \gamma A_i,$$

y los residuos correspondientes vienen dados por

$$e_i = Y_i - \hat{Y}_i = Y_i - (\alpha + \beta P_i + \gamma A_i).$$

El análisis de regresión asigna valores a los parámetros del modelo ($\alpha$, $\beta$ y $\gamma$), de manera que las $\hat{Y}_i$ sean tan parecidas a las $Y_i$ como sea posible. Esto se logra eligiendo valores que minimicen la suma de los cuadrados de los residuos, lo que conduce al apelativo de *mínimos cuadrados ordinarios* (MCO) para las estimaciones resultantes.[6] Cuando se ejecuta esta minimización en una muestra concreta se dice que se procede a la *estimación* de los parámetros de la regresión. En ocasiones se dice que los maestros de la econometría que estiman modelos de regresión a diario se dedican a «hacer regresiones», aunque muchas veces parece que son las regresiones las que nos hacen a nosotros, y no al contrario. El apéndice de este capítulo bosqueja los aspectos formales de la estimación por regresión y la teoría estadística subyacente.

Al hacer la regresión (2.1) sobre los datos de los cinco estudiantes de los grupos A y B se obtienen las estimaciones siguientes (estas estimaciones pueden realizarse con una calculadora de bolsillo, pero en el trabajo empírico real se utilizan programas profesionales de regresión):

$$\alpha = 40.000$$
$$\beta = 10.000$$
$$\gamma = 60.000.$$

[6] Aquí el adjetivo *ordinarios* se refiere a que el procedimiento asigna el mismo peso a la hora de efectuar la suma de los cuadrados. La estimación por medio de mínimos cuadrados ponderados se trata en el capítulo 5.

El coeficiente de centros privados es en este caso 10.000, lo que implica un diferencial de ingresos privado-público de 10.000 dólares. Esta cantidad es en realidad un promedio ponderado de los dos efectos específicos de grupo (recordemos que el efecto en el grupo A es de –5.000 y en el grupo B de 30.000). El resultado no es ni la media simple (12.500), ni la media ponderada según el tamaño de los grupos (9.000), pero no cae demasiado lejos de ninguna de las dos. En este caso la regresión asigna un peso de 4/7 al grupo A y otro de 3/7 al B. Como en el caso de las otras medias ponderadas, el promedio ponderado que da la regresión es bastante inferior a la diferencia de ingresos que daría la comparación sin controlar entre alumnos de escuelas privadas y públicas.[7]

La estimación por medio de regresión (y los correspondientes errores típicos para calcular la varianza muestral) se efectúa de manera inmediata mediante computadoras y programas econométricos. La simplicidad computacional y la interpretación conceptual de las estimaciones por regresión como un promedio ponderado de las diferencias dentro de cada grupo son dos de las razones por las que recurrimos a esta herramienta. La regresión tiene otras dos cualidades que la hacen atractiva. Primero, es costumbre entre los maestros aportar estimaciones por regresión en cualquier investigación econométrica sobre efectos causales, incluidas las que involucren variables de tratamiento que adopten más de dos valores. La estimación por regresión proporciona una base simple sobre la que se pueden aplicar técnicas más refinadas. En segundo lugar, en ciertas circunstancias las estimaciones por regresión son efectivas en el sentido de que proporcionan las estimaciones estadísticas de los efectos causales promedio más precisas a las que se puede aspirar a partir de una determinada muestra. Este aspecto técnico se trata brevemente en el apéndice de este capítulo.

### *Lo público y lo privado cara a cara*

El banco de datos C&B incluye más de 14.000 antiguos estudiantes. Estos estudiantes fueron admitidos y rechazados en muchas combi-

[7] En nuestro libro *Mostly Harmless Econometrics* (*Econometría casi inocua*, Princeton University Press, 2009) tratamos en más detalle los sistemas de pesos en regresiones.

naciones diferentes de universidades (C&B pedía los nombres de al menos tres centros que los estudiantes hubieran considerado seriamente, aparte de aquél en el que estudiaron). Muchos de los conjuntos posibles de solicitud/admisión en estos datos están representados por un solo estudiante. Además, muchos conjuntos con más de un estudiante están formados por centros que son todos ellos, o bien públicos, o bien privados. Como en los grupos C y D de la tabla 2.1, estos grupos totalmente homogéneos no arrojan luz sobre el valor de la formación en centros privados.

Se puede aumentar el número de comparaciones útiles si se consideran equivalentes centros que sean igual de selectivos en cuanto a criterios de acceso, en lugar de insistir en que se trate exactamente de los mismos centros. Con el fin de agrandar los grupos que se obtienen de esta manera, consideraremos comparables dos centros si caen dentro de la misma categoría selectiva de Barron's.[8] Volviendo a nuestra matriz de emparejamiento simplificada, supongamos que Omniópolis y Altópolis estuvieran en la categoría «competitiva», Alterópolis y Astutosa en «altamente competitiva», y que Treposa y Foliosa fueran «de las más competitivas». Según el escalafón de Barron's, toda persona que solicitara ingresar en Altópolis, Astutópolis y Foliosa, y fuera admitida en Altópolis y Astutosa, se podría comparar con un estudiante que pidiera Omniópolis, Astutosa y Treposa, pero que resultara admitido en Omniópolis y Astutosa. Los estudiantes de ambos grupos solicitaron ingreso en una universidad «competitiva», una «altamente competitiva» y una «de las más competitivas», y fueron admitidos en una «competitiva» y en otra «altamente competitiva».

Los datos C&B permiten emparejar de este modo a 9.202 estudiantes. Pero como lo que nos interesa es la comparación entre centros públicos y privados, nuestra muestra construida a partir de los criterios de Barron's debe restringirse a grupos concordes que contengan estudiantes de centros de los dos tipos. Esto deja 5.583 estudiantes emparejados para el análisis, distribuidos entre 151 grupos de

[8] Barron's clasifica las facultades en grupos según el percentil en que caen las calificaciones de los estudiantes matriculados, y en función de la proporción de admisiones frente a solicitudes. Los grupos son: «de las más competitivas», «altamente competitiva», «muy competitiva», «competitiva», «menos competitiva» y «no competitiva».

semejanza por su grado de selectividad que contienen alumnado de universidades tanto públicas como privadas.

Nuestro modelo operativo de regresión para la muestra construida a partir de los criterios de Barron's difiere en varios aspectos de la regresión mostrada en la ecuación (2.1), empleada para analizar la matriz de emparejamiento de la tabla 2.1. En primer lugar, el modelo operativo coloca en el primer miembro el logaritmo natural de los ingresos, en lugar de usar los ingresos directamente. Como se explica en el apéndice de este capítulo, usar una variable dependiente de carácter logarítmico permite interpretar las estimaciones de la regresión como cambios porcentuales. Por ejemplo, si se estima para $\beta$ un valor 0,05, entonces los alumnos de escuelas privadas ganan aproximadamente un 5% más que los de escuelas públicas, resultado condicionado a los controles que se hayan incluido en el modelo.

Otra diferencia importante entre nuestro modelo empírico y el ejemplo de la tabla 2.1 consiste en que ahora incluimos más variables de control, mientras que en el ejemplo sólo consta la variable binaria $A_i$, que identifica a los estudiantes del grupo A. Los controles clave en el modelo actual son un conjunto de muchas variables binarias que identifican todos los emparejamientos según los criterios de Barron's representados en la muestra (dejando fuera un grupo, como categoría de referencia). Estos controles representan la selectividad relativa de las universidades solicitadas por los estudiantes, y las de los centros que los admitieron que, en el mundo real, representan muchas combinaciones de centros. El modelo de regresión resultante tiene esta pinta:

$$\text{en } Y_i = \alpha + \beta P_i + \sum_{j=1}^{150} \gamma_j GROUP\, j_i + \delta_1 SAT_i + \delta_2 \text{ en } PI_i + e_i \qquad (2.2)$$

El parámetro $\beta$ de este modelo sigue representando el efecto del tratamiento de interés, una estimación del efecto causal de estudiar en un centro privado. Pero este modelo contiene 151 grupos de control, en lugar de los dos presentes en el ejemplo. Los parámetros $\gamma_j$, para $j = 1,\ldots,150$, son los coeficientes de las variables binarias de los 150 grupos de selección, denotadas como $GROUP_{ji}$.

Vale la pena analizar algo más la notación de la ecuación (2.2),

porque volveremos a utilizarla más adelante. La variable binaria $GROUP_{ji}$ vale 1 cuando el estudiante $i$ pertenece al grupo $j$, y vale 0 en los demás casos. Por ejemplo, la primera variable binaria, denotada como $GROUP_{1i}$, podría identificar a los estudiantes que solicitaron el ingreso y fueron admitidos en dos universidades «altamente competitivas». La segunda, $GROUP_{2i}$, podría identificar a los sujetos que solicitaron dos facultades de la categoría «altamente competitiva» y una «de las más competitivas», pero fueron admitidos en una de cada clase. No es relevante el orden en el que se codifican las categorías, mientras haya una variable binaria para cada combinación posible, omitiendo siempre un grupo como referencia. Aunque hayamos pasado de una variable binaria a 150, la idea es la misma de antes: controlar los conjuntos de centros de solicitud y admisión nos hace avanzar un paso gigantesco en pos de la comparación en condiciones *ceteris paribus* entre estudiantes de universidades públicas y privadas.

Se añaden dos variables de control adicionales como modificación final con fines operativos: la puntuación individual en las pruebas de acceso a la universidad $(SAT_i)$ y el logaritmo de los ingresos de sus padres $(PI_i)$, aparte de otras variables diversas cuya explicación relegamos a una nota al pie.[9] Los controles de la calificación individual en las pruebas de acceso y del logaritmo de los ingresos paternos aparecen en el modelo como los coeficientes $\delta_1$ y $\delta_2$ (léanse «delta-1» y «delta-2»), respectivamente. Este tipo de controles sirven como medidas directas de las aptitudes individuales (notas de acceso) y del contexto familiar (ingresos de los padres), y pueden ayudar a que la comparación entre centros públicos y privados junte manzanas con manzanas y naranjas con naranjas mejor de lo que lo haría si no se introdujeran. A la vez, dependiendo de cómo funcionen las variables binarias que marcan los distintos grupos, podría ocurrir que algunos de estos controles no fueran necesarios, un aspecto que se tratará en detalle más adelante.

[9] Otros controles del modelo empírico incluyen variables binarias que marcan según género, raza, aptitudes deportivas, y estudiantes que obtuvieron el grado universitario dentro del 10% mejor de su clase. Estas variables no constan en la ecuación (2.2).

### *Regresiones en acción*

Empezamos estimando la mejora de ingresos que resulta de acudir a un centro privado aplicando un modelo estadístico carente de controles. Al hacer la regresión del logaritmo de los ingresos (1995) con la variable binaria que indica si se estudió en un centro público, sin incluir en el modelo otros regresores (variables del segundo miembro), se obtiene la diferencia bruta del logaritmo de los ingresos entre quienes acudieron a universidades públicas y el resto de la muestra (el apéndice del capítulo explica por qué al usar una sola variable binaria se extrae la diferencia entre las medias de los dos grupos definidos por esa variable). No debe sorprender que esta diferencia bruta, que consta en la primera columna de la tabla 2.2, muestre una ventaja sustancial para las universidades privadas. En concreto, se estima que quienes estudian en universidades privadas ganan un 14% más que el resto.

Los números entre paréntesis bajo las estimaciones de la regresión de la tabla 2.2 son los errores típicos estimados para esos valores. Al igual que los errores típicos de una diferencia de medias que se explican en el apéndice del capítulo 1, estos errores típicos cuantifican la precisión estadística de las estimaciones que arroja la regresión. El error típico asociado a la cantidad estimada en la columna (1) asciende a 0,055. El hecho de que 0,135 sea más de dos veces el error típico (0,055) hace muy poco probable que la diferencia positiva estimada entre centros privados y públicos sea resultado del puro azar. El coeficiente de los centros privados es estadísticamente significativo.

La gran diferencia a favor de los centros privados que aparece en la columna (1) de la tabla 2.2 constituye un hecho descriptivo interesante pero, como en nuestro ejemplo anterior, sin duda parte de este efecto se debe a un sesgo de selección. Como veremos más adelante, los estudiantes de universidades privadas tienen notas mejores en las pruebas de acceso y proceden de familias con más recursos que los estudiantes de las universidades públicas, y por eso cabe esperar que ganen más dinero con independencia de dónde cursaran estudios. Por lo tanto, introducimos controles que tienen en cuenta las capacidades individuales y el entorno familiar a la hora de estimar la ventaja de acudir a un centro privado. La columna (2) de la tabla 2.2 presenta una estimación de la ventaja de las universidades privadas a partir

## Tabla 2.2. Efectos de la universidad privada: emparejamientos de Barron's

| | Sin controles de selección | | | Con controles de selección | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Centro privado | .135 | .095 | .086 | .007 | .003 | .013 |
| | (.055) | (.052) | (.034) | (.038) | (.039) | (.025) |
| Nota de acceso individual ÷ 100 | | .048 | .016 | | .033 | .001 |
| | | (.009) | (.007) | | (.007) | (.007) |
| Logaritmo de los ingresos paternos | | | .219 | | | .190 |
| | | | (.022) | | | (.023) |
| Mujeres | | | –.403 | | | –.395 |
| | | | (.018) | | | (.021) |
| Negros | | | .005 | | | –.040 |
| | | | (.041) | | | (.042) |
| Hispanos | | | .062 | | | .032 |
| | | | (.072) | | | (.070) |
| Asiáticos | | | .170 | | | .145 |
| | | | (.074) | | | (.068) |
| Otra raza (o no consta) | | | –.074 | | | –.079 |
| | | | (.157) | | | (.156) |
| En el 10% mejor de su escuela secundaria | | | .095 | | | .082 |
| | | | (.027) | | | (.028) |
| No consta nivel de secundaria | | | .019 | | | .015 |
| | | | (.033) | | | (.037) |
| Deportista | | | .123 | | | .115 |
| | | | (.025) | | | (.027) |
| Variables binarias de selección | No | No | No | Sí | Sí | Sí |

*Notas:* Esta tabla incluye la estimación de los efectos que ejerce sobre los ingresos el hecho de cursar estudios en centros universitarios privados. Cada columna muestra los coeficientes de una regresión del logaritmo de los ingresos con variables binarias que marcan si se estudió en un centro privado, así como varios controles. Los resultados de las columnas (4)-(6) proceden de modelos que incluyen variables binarias para diversos grupos de centros según su nivel selectivo. El tamaño de la muestra es de 5.583. Los errores típicos constan entre paréntesis.

de un modelo que incluye controles sobre las notas en las pruebas de acceso. A cada 100 puntos en las pruebas de acceso les corresponde un incremento del 5% en los ingresos. Al controlar las calificaciones de acceso, la ventaja de los centros privados se reduce a aproximadamente 0,1. Si se añaden controles relacionados con los ingresos paternos, así como con características demográficas relacionadas con raza o género, estar entre los mejores de la clase o si el estudiante es atleta,[10] se rebaja un poco más la ventaja de los centros privados, hasta un valor aún considerable y estadísticamente significativo igual a 0,086, que consta en la columna (3) de la tabla.

Aunque se trate de una cifra sustancial, probablemente aún sea demasiado elevada, es decir, sigue contaminada por efectos de selección. La columna (4) consigna las estimaciones que resultan cuando no se controlan las capacidades individuales, el contexto familiar o las características demográficas. Pero observemos, en cambio, que el modelo de regresión empleado para calcular las estimaciones que figuran en esta columna incluyen una variable binaria que etiqueta cada grupo selectivo de universidades de la muestra. Es decir, el modelo que se usa para construir esta estimación incorpora la variable binaria $GROUP_{ji}$, con $j$ =1,…, 150 (la tabla omite la multitud de valores $\gamma_j$ que produce este modelo, pero indica su inclusión en la fila denominada «controles de selección»). La ventaja que se estima por haber estudiado en centros privados cuando se incluyen los controles por grupo selectivo coincide casi exactamente con un valor cero, con un error típico en torno a 0,04. Y eso no es todo: tras aniquilar la ventaja de las universidades privadas con las variables binarias de grupos selectivos, las columnas (5) a (6) muestran que la ventaja residual varía poco cuando se introducen en el modelo controles de capacidad individual o de contexto familiar. Esto parece indicar que los controles por solicitudes y por admisiones nos acercan a comparar manzanas con manzanas y naranjas con naranjas que, como sabemos, es la base de cualquier estrategia creíble de regresión en pos de la inferencia causal.

Los resultados de las columnas (4) a (6) de la tabla 2.2 proceden de la submuestra de 5.583 estudiantes para los que se pueden

[10] En EE. UU., los jóvenes con dotes atléticas especiales suelen tener grandes facilidades para acceder a las universidades, tanto públicas como privadas. *(N. de la T.)*

construir emparejamientos basados en los criterios de Barron's que permite hacer comparaciones dentro de cada grupo entre alumnos de universidades públicas y privadas. Quizá haya algo especial en esta submuestra, que contiene menos de la mitad del total de encuestados en C&B. Esta posibilidad sugiere probar un sistema de controles menos exigente, que incluya sólo la nota media en las pruebas de acceso a la universidad a la hora de definir los grupos de universidades en las que se presenta la solicitud, más otra variable binaria para el número de centros solicitados (es decir, una variable binaria que marca a los estudiantes que pidieron el ingreso en dos universidades, otra para los que lo hicieron en tres, etcétera), en vez de todo un conjunto de 150 variables binarias por grupos selectivos. Esta regresión, que puede aplicarse a la totalidad de los datos C&B, recibe el nombre de *modelo de autorrevelado*, porque encuentra justificación en el hecho de que son los propios solicitantes quienes mejor conocen sus capacidades y dónde es probable que sean admitidos. Esta autoevaluación se refleja en el número y en la selectividad media de las universidades que se solicitan. Por regla general, los solicitantes más flojos cursan menos solicitudes, y lo hacen en universidades menos selectivas, que los estudiantes mejor dotados.

El modelo de autorrevelado genera resultados muy semejantes a los que resultan de los emparejamientos según los criterios de Barron's. Las estimaciones del autorrevelado, calculadas a partir de una muestra de 14.238 sujetos, se presentan en la tabla 2.3. Como antes, las tres primeras columnas de la tabla evidencian una caída notable de la ventaja de los centros privados, aunque se mantenga en un nivel sustancial, a medida que se añaden controles para la capacidad individual y el contexto familiar (en este caso cae de 0,21 hasta 0,14). Al mismo tiempo, las columnas (4) a (6) revelan que los modelos con controles para el número de solicitudes y la selectividad promedio de las universidades solicitadas arrojan resultados estadísticamente no significativos en el entorno de 0,03. Además, como en los modelos que controlan según los grupos de Barron's, los modelos con controles basados en la selectividad promedio generan estimaciones muy poco sensibles a la inclusión de controles sobre la capacidad individual o el contexto familiar.

La asistencia a una universidad privada parece, por lo tanto, no guardar relación con los ingresos futuros una vez se tiene en cuenta

## Tabla 2.3. Efectos de la universidad privada: controles según la nota media en las pruebas de acceso

| | Sin controles de selección | | | Con controles de selección | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Centro privado | .212 | .152 | .139 | .024 | .031 | .037 |
| | (.060) | (.057) | (.043) | (.062) | (.062) | (.039) |
| Nota de acceso individual ÷ 100 | | .051 | .024 | | .036 | .009 |
| | | (.008) | (.006) | | (.006) | (.006) |
| Logaritmo de los ingresos paternos | | | .181 | | | .159 |
| | | | (.026) | | | (.025) |
| Mujeres | | | –.398 | | | –.396 |
| | | | (.012) | | | (.014) |
| Negros | | | –.003 | | | –.037 |
| | | | (.031) | | | (.035) |
| Hispanos | | | .027 | | | .001 |
| | | | (.052) | | | (.054) |
| Asiáticos | | | .189 | | | .155 |
| | | | (.035) | | | (.037) |
| Otra raza (o no consta) | | | –.166 | | | –.189 |
| | | | (.118) | | | (.117) |
| En el 10% mejor de su escuela secundaria | | | .067 | | | .064 |
| | | | (.020) | | | (.020) |
| No consta nivel de secundaria | | | .003 | | | -.008 |
| | | | (.025) | | | (.023) |
| Deportista | | | .107 | | | .092 |
| | | | (.027) | | | (.024) |
| Puntuación media en pruebas de acceso en las escuelas solicitadas ÷ 100 | | | | .110 | .082 | .077 |
| | | | | (.024) | (.022) | (.012) |
| Presentaron dos solicitudes | | | | .017 | .062 | .058 |
| | | | | (.013) | (.011) | (.010) |
| Presentaron tres solicitudes | | | | .093 | .079 | .066 |
| | | | | (.021) | (.019) | (.017) |
| Presentaron cuatro o más solicitudes | | | | .139 | .127 | .098 |
| | | | | (.024) | (.023) | (.020) |

*Notas:* Esta tabla incluye la estimación de los efectos que ejerce sobre los ingresos el hecho de cursar estudios en centros universitarios privados. Cada columna muestra los coeficientes de una regresión del logaritmo de los ingresos con variables binarias que marcan si se estudió en un centro privado, así como varios controles. El tamaño de la muestra es 14.238. Los errores típicos constan entre paréntesis.

## Tabla 2.4. Efectos del carácter selectivo de los centros: controles según la nota media en las pruebas de acceso

| | Sin controles de selección | | | Con controles de selección | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Nota media de acceso al centro ÷ 100 | .109 | .071 | .076 | -.021 | -.031 | .000 |
| | (.026) | (.025) | (.016) | (.026) | (.026) | (.018) |
| Nota de acceso individual ÷ 100 | | .049 | .015 | | .037 | .009 |
| | | (.007) | (.006) | | (.006) | (.006) |
| Logaritmo de los ingresos paternos | | | .187 | | | .161 |
| | | | (.024) | | | (.025) |
| Mujeres | | | –.403 | | | –.396 |
| | | | (.015) | | | (.014) |
| Negros | | | –.023 | | | –.034 |
| | | | (.035) | | | (.035) |
| Hispanos | | | .015 | | | .006 |
| | | | (.052) | | | (.053) |
| Asiáticos | | | .173 | | | .155 |
| | | | (.036) | | | (.037) |
| Otra raza (o no consta) | | | –.188 | | | –.193 |
| | | | (.119) | | | (.116) |
| En el 10% mejor de su escuela secundaria | | | .061 | | | .063 |
| | | | (.018) | | | (.019) |
| No consta nivel de secundaria | | | .001 | | | -.009 |
| | | | (.024) | | | (.022) |
| Deportista | | | .102 | | | .094 |
| | | | (.025) | | | (.024) |
| Puntuación media en pruebas de acceso en las escuelas solicitadas ÷ 100 | | | | .138 | .016 | .089 |
| | | | | (.017) | (.015) | (.014) |
| Presentaron dos solicitudes | | | | .082 | .075 | .063 |
| | | | | (.015) | (.014) | (.011) |
| Presentaron tres solicitudes | | | | .107 | .096 | .074 |
| | | | | (.026) | (.024) | (.022) |
| Presentaron cuatro o más solicitudes | | | | .153 | .143 | .106 |
| | | | | (.031) | (.030) | (.025) |

*Notas:* Esta tabla incluye la estimación de los efectos que ejerce sobre los ingresos el carácter más o menos selectivo de la universidad. Cada columna muestra los coeficientes de una regresión del logaritmo de los ingresos con la nota media de acceso del centro donde se cursan los estudios, así como varios controles. El tamaño de la muestra es 14.238. Los errores típicos constan entre paréntesis.

el sesgo de selección. Pero quizá nos estamos equivocando al poner el foco en la comparación entre lo público y lo privado. Los estudiantes podrían beneficiarse de asistir a las universidades Treposa, Foliosa o Astutosa por el simple hecho de que sus compañeros de clase, en esos centros, son mucho mejores. La sinergia que surge en un grupo de compañeros más capaces podría ser el rasgo por el que valga la pena pagar la factura de un centro privado.

Podemos explorar esta hipótesis si reemplazamos la variable binaria de centro privado en el modelo de autorrevelado por alguna medida de la calidad de los compañeros. En concreto, y como en el estudio original de Dale y Krueger que inspira nuestro análisis, reemplazamos $P_i$ en la ecuación (2.2) por la puntuación media en las pruebas de acceso de los compañeros de la universidad en la que se cursaron los estudios.[11] Las columnas (1) a (3) de la tabla 2.4 revelan que los estudiantes que acudieron a universidades más selectivas tienen un éxito marcadamente superior en el mercado de trabajo, con un efecto positivo del carácter selectivo de los centros del 8% en los ingresos por cada 100 puntos de incremento medio en la nota de corte para ser admitido. Aun así, este efecto también parece ser un artificio inducido por el sesgo de selección, debido a la mayor ambición y mayores capacidades de quienes estudian en centros selectivos. Las estimaciones de modelos con controles de autorrevelado, que constan en las columnas (4) a (6) de la tabla, muestran que el carácter selectivo de la universidad carece, esencialmente, de relación con los ingresos.

## 2.3 *¿Ceteris paribus*?

Tema: Describa brevemente las experiencias, retos y logros que lo definen a usted como persona.

Redacción: Tengo una personalidad dinámica; se me suele ver escalando montes y glaciares. Consigo cocinar las recetas de bizcocho de treinta minutos en tan sólo veinte. Soy un experto estucador, amante veterano y proscrito en Perú. Los miércoles, después de las clases, reparo electrodomésticos por amor al arte.

[11] Dale y Krueger, «Estimating the Payoff to Attending a More Selective College», *Quarterly Journal of Economics*, 2002.

> Soy artista abstracto, analista concreto y un despiadado ratón de biblioteca. Vibrante, driblador, inquieto, pero pago mis cuentas. Triunfé en los toros en San Juan, en las competiciones de buceo en arrecifes de Sri Lanka y en los concursos de deletreo del Kremlin. He interpretado a Hamlet, he realizado operaciones quirúrgicas a corazón abierto y he hablado con Elvis.
> Pero todavía no he ido a la universidad.
>
> De una redacción compuesta por Hugh Gallagher, de 19 años
> (Hugh asistió luego a la Universidad de Nueva York)

Imagine a Harvey y Uma cuando recibieron sus cartas de admisión. Ambos están encantados de haber entrado en Harvard (esto tiene que ser como uno de esos bizcochos hechos en veinte minutos). Harvey de inmediato acepta la oferta de Harvard. ¿Y quién no? Pues Uma, que se enfrenta a una decisión difícil y elige U-Mass. ¿Qué le pasa a Uma? ¿De verdad sus *ceteris* son *paribus*?

Uma podría tener buenos motivos para elegir U-Mass antes que Harvard, a pesar de su menor prestigio. El precio es una circunstancia obvia (Uma consiguió una de las becas Adams de Massachusetts, que cubre la matrícula universitaria de estudiantes buenos como ella, pero que no se puede usar en centros privados). Si el precio le importa a Uma más que a Harvey, entonces es posible que las circunstancias de Uma difieran de las de Harvey en muchos otros aspectos. Quizá sea más pobre. Algunos de nuestros controles de regresión tienen en cuenta los ingresos paternos, pero esta es una medida imperfecta del nivel de vida de una familia. Entre otras cosas, ignoramos cuántos hermanos y hermanas tenían los estudiantes de la muestra C&B. Una familia mayor puede tener más difícil costear la educación de toda la prole con el mismo nivel de ingresos. Si el tamaño de la familia estuviera también relacionado con los ingresos posteriores (véase más sobre esto en el capítulo 3), entonces nuestras estimaciones mediante regresión de las ventajas de los centros privados no estarían comparando manzanas con manzanas después de todo.

Esto es algo más que una historia para contar ante un fuego de campamento. La regresión pretende ser un modo de mantener el resto de las cosas iguales, pero la igualdad se genera sólo para las variables incluidas en el segundo miembro del modelo. No introdu-

cir suficientes controles, o los adecuados, abre la puerta al sesgo de selección. La versión del sesgo de selección que aparece en las regresiones cuando los controles no son los adecuados se denomina *sesgo de variables omitidas* (SVO), y constituye una de las ideas centrales del canon de la econometría.

Volvamos a nuestro ejemplo de cinco estudiantes para ilustrar el SVO y la influencia de omitir el control sobre la pertenencia al grupo A de solicitantes. Aquí la «regresión larga» incluye la variable binaria $A_i$, que marca a los sujetos del grupo A. El modelo de regresión que incluye esta variable se escribe como

$$Y_i = \alpha^l + \beta^l P_i + \gamma A_i + e_i^l. \tag{2.3}$$

Se trata de la ecuación (2.1) reescrita con el superíndice $l$ en los parámetros y los residuos para recordarnos que la ordenada en el origen y el coeficiente de centro privado corresponden al modelo «largo», y para facilitar la comparación con el modelo simple que pasamos a plantear.

¿Influye la inclusión de $A_i$ sobre la estimación del efecto de estudiar en centros privados en la regresión anterior? Supongamos que procedemos a una regresión simple, sin controles, que podría expresarse como

$$Y_i = \alpha^s + \beta^s P_i + \gamma A_i + e_i^s.$$

Como el único regresor aquí es una variable binaria, la pendiente de este modelo resultará igual a la diferencia del $Y_i$ promedio entre los sujetos que tienen $P_i$ activado y los que lo tienen igual a cero. Como indicamos en el apartado 2.1, $\beta^s = 20.000$ en la regresión simple, mientras que el parámetro de la regresión larga, $\beta^l$, sólo asciende a 10.000. La diferencia entre $\beta^l$ y $\beta^s$ es el SVO debido a la omisión de $A_i$ en la regresión simple. En este caso el SVO asciende a 10.000 dólares, una cifra de la que vale la pena preocuparse.

¿Por qué es tan intenso el efecto de omitir el control del grupo A sobre la estimación del efecto de estudiar en un centro privado? Recordemos que los ingresos medios de los estudiantes del grupo A superan los del grupo B. Además, dos tercios de los estudiantes de este grupo de ingresos elevados acudió a centros privados, mientras

que sólo lo hizo la mitad de los integrantes del grupo B, con ingresos inferiores. La diferencia de ingresos entre estudiantes de centros públicos y privados se debe en parte al hecho de que los estudiantes del grupo A, en su mayoría de centros privados, tienen ingresos superiores de todos modos, con independencia de esa circunstancia. Al incluir una variable binaria que controla la pertenencia al grupo A en la regresión larga se tiene en cuenta esta diferencia.

Como sugiere esta explicación, la conexión formal entre los coeficientes de las regresiones simple y larga tiene dos componentes:

*(i)* La relación entre la variable omitida $(A_i)$ y la variable de tratamiento $(P_i)$; pronto veremos cómo cuantificar esta relación por medio de una regresión adicional.
*(ii)* La relación entre la variable omitida $(A_i)$ y la variable de resultado $(Y_i)$. Esta relación viene dada por el coeficiente de la variable omitida en la regresión larga, en este caso el parámetro $\gamma$ de la ecuación (2.3).

Al reunir estas piezas se obtiene la *fórmula del SVO*. Partimos del hecho de que

$$\begin{aligned}&\textit{Efecto de } P_i \textit{ en simple} = \textit{Efecto de } P_i \textit{ en larga} + \\ &+ ([\textit{Relación entre omitida e incluida}] \times \\ &\times [\textit{Efecto de la omitida en larga}]).\end{aligned}$$

En concreto, cuando la variable omitida es $A_i$ y la variable de tratamiento $P_i$, entonces

$$\begin{aligned}&\textit{Efecto de } P_i \textit{ en simple} = \textit{Efecto de } P_i \textit{ en larga} + \\ &+ ([\textit{Relación entre } A_i \textit{ y } P_i] \times \\ &\times [\textit{Efecto de } A_i \textit{ en larga}]).\end{aligned}$$

El sesgo de variables omitidas (SVO), definido como la diferencia entre los coeficientes de $P_i$ en los modelos simple y largo, consiste en una sencilla reordenación de la ecuación anterior:

$$\begin{aligned}&\textit{SVO} = [\textit{Relación entre } A_i \textit{ y } P_i] \times \\ &\times [\textit{Efecto de } A_i \textit{ en larga}].\end{aligned}$$

Podemos refinar la fórmula del SVO si recurrimos al hecho de que los dos términos de la fórmula son en sí mismos coeficientes de regresión. El primer término es el coeficiente que resulta de calcular una regresión de la variable omitida $A_i$ con la variable binaria de centro privado. En otras palabras, este término es el coeficiente $\varpi_1$ (léase «pi-uno») en el modelo de regresión

$$A_i = \pi_0 + \pi_1 P_i + u_i,$$

donde $u_i$ es un residuo. Ahora podemos escribir la fórmula del SVO de manera compacta con letras griegas:

$$\begin{aligned} SVO &= \textit{Efecto de } P_i \textit{ en simple} - \textit{Efecto de } P_i \textit{ en larga} \\ &= \beta^s - \beta^l = \pi_1 \times \gamma, \end{aligned}$$

donde $\gamma$ es el coeficiente de $A_i$ en la regresión larga. Esta fórmula, muy importante, se deduce en el apéndice del presente capítulo.

El alumnado de centros privados incluye dos sujetos del grupo A y dos sujetos del grupo B, mientras que el colectivo que estudió en centros públicos tiene a una persona en A y otra en B. El coeficiente $\varpi_1$ de nuestro ejemplo con cinco estudiantes vale, por tanto, $2/3 - 1/2 = 0{,}1667$. Como se indica en el apartado 2.2, el coeficiente $\gamma$ vale 60.000, lo que refleja que el grupo A tiene ingresos superiores. Si se reúnen las piezas tenemos que

$$\begin{aligned} SVO &= \textit{Simple} - \textit{Larga} \\ &= \beta^s - \beta^l \\ &= 20.000 - 10.000 = 10.000 \end{aligned}$$

y que

$$\begin{aligned} SVO = [&\textit{Regresión de omitida sobre incluida}] \times \\ &\times [\textit{Efecto de la omitida en larga}] = \\ &= \pi_1 \times \gamma = 0{,}1667 \times 60.000 = 10.000. \end{aligned}$$

¡Caramba! El cálculo sugerido por la fórmula del SVO coincide de verdad con la comparación directa de los coeficientes resultantes de las regresiones simple y larga.

La fórmula del SVO es un resultado matemático que explica las diferencias entre coeficientes de regresión cuando se comparan dos escenarios cualesquiera de tipos simple y largo, con independencia de la interpretación causal de los parámetros de la regresión. Las etiquetas «simple» y «larga» son tan sólo relativas: la regresión simple no tiene por qué ser especialmente simple, pero la regresión larga siempre es más larga, porque incluye los mismos regresores que la simple y al menos uno más. Con frecuencia las variables adicionales que alargan la regresión larga son de carácter hipotético, es decir, no están disponibles en los datos. La fórmula del SVO es una herramienta que permite considerar el impacto de las variables de control de las que *querríamos* disponer. Esto, a su vez, ayuda a aclarar si realmente *ceteris* es o no es *paribus*. Lo que nos lleva de nuevo a Uma y Harvey.

Supongamos que una de las variables omitidas en la ecuación (2.2) fuera el tamaño de la unidad familiar, $FS_i$. Hemos incluido los ingresos paternos como variable de control, pero no el número de hermanos y hermanas que podrían acudir también a la universidad, una información que no está disponible en el banco de datos C&B. Si la variable omitida es $FS_i$, tenemos

$$\begin{aligned} SVO &= Simple - \text{Larga} = \\ &= [Relación\ entre\ FS_i\ y\ P_i] \times \\ &\quad \times [Efecto\ de\ FS_i\ en\ larga]. \end{aligned}$$

¿Por qué la omisión del tamaño de las familias puede sesgar las estimaciones por regresión del efecto de estudiar en centros privados? Porque las diferencias de ingresos entre los graduados de Harvard y de U-Mass provienen en parte de diferencias en cuanto a los tamaños de las familias entre los dos grupos (es decir, de la relación entre $FS_i$ y $P_i$) y del hecho de que las familias más reducidas están correlacionadas con ingresos superiores, incluso después de introducir las variables de control presentes en la regresión simple (este es el efecto de $FS_i$ sobre la regresión larga, que incluye igualmente todos los controles anteriores). La regresión larga tiene en cuenta el hecho de que los estudiantes que acuden a Harvard proceden de familias (en promedio) más pequeñas que las de los estudiantes que fueron a U-Mass, mientras que la regresión simple que omite $FS_i$ no lo tiene en cuenta.

El primer término en esta aplicación de la fórmula del SVO es el coeficiente que resulta de una regresión entre la variable omitida ($FS_i$) y la incluida ($P_i$) y todo lo demás que aparezca en el segundo miembro de la ecuación (2.2). Esta regresión, que a veces recibe el nombre de «auxiliar» porque facilita la interpretación de la regresión que de verdad nos importa, se puede escribir como

$$FS_i = \pi_0 + \pi_1 P_i + \sum_j \theta_j GROUP_{ji} + \pi_2 SAT_i + \pi_3 \text{ en } PI_i + \mu_i. \qquad (2.4)$$

La mayor parte de los coeficientes de la ecuación (2.4) tienen poco interés. Lo que importa aquí es $\pi_1$, porque mide la relación entre la variable omitida, $FS_i$, y la variable que representa el efecto que queremos valorar, $P_i$, tras tener en cuenta el resto de variables que aparecen tanto en el modelo simple como en el largo.[12]

Para completar la fórmula del SVO en este caso escribiríamos la regresión larga como:

$$\begin{aligned} \text{en } Y_i = \alpha^l + \beta^l P_i + \sum_j \gamma_j^l GROUP_{ji} + \\ + \delta_1^l SAT_i + \delta_2^l \text{ en } PI_i + \lambda FS_i + e_i^l, \end{aligned} \qquad (2.5)$$

donde de nuevo se emplea el superíndice $l$ para indicar la regresión «larga». El regresor $FS_i$ figura aquí con el coeficiente $\lambda$.[13] Así, la fórmula del SVO queda

$$SVO = Simple - Larga = \beta - \beta_l = \pi_1 \times \lambda,$$

donde $\beta$ procede de la ecuación (2.2).

Si seguimos pensando en la ecuación (2.2) como la regresión simple, mientras que la larga incorpora ahora todas las variables incluidas ahí más el tamaño de la familia, vemos que es probable que el SVO sea positivo. Los estudiantes de centros privados tienden a formar parte, en promedio, de familias más pequeñas, incluso teniendo en cuenta los ingresos familiares. Si esto es así, el coeficiente de regresión que relaciona el tamaño familiar con la asistencia a

[12] Las variables binarias de grupo en (2.4), $\theta_j$, se leen «zeta-jota».
[13] Este coeficiente se lee «lambda».

centros privados es negativo ($\pi_1 < 0$ en la ecuación [2.4]). Los estudiantes procedentes de familias menores tienen también más posibilidades de lograr más ingresos, con independencia del tipo de centro donde estudiaran, así que el efecto de omitir el control sobre el tamaño familiar en una regresión larga también será negativo ($\lambda < 0$ en la ecuación [2.5]). El producto de estos dos términos negativos es positivo.

Razonar con cuidado acerca del SVO es una parte esencial del juego econométrico. No podemos usar datos que cuantifiquen las consecuencias de omitir variables que no se observan, pero podemos emplear la fórmula del SVO para emitir una conjetura bien fundamentada acerca de las consecuencias probables de su omisión. La mayoría de las variables de control que pueden omitirse en la ecuación (2.2) se asemejan al tamaño familiar en que el signo del SVO que resulta al omitirlas probablemente sea positivo. Deducimos de ahí que, a pesar de lo reducido de los efectos de asistir a un centro privado que se muestran en las columnas (4)-(6) de las tablas 2.2 y 2.3, bien pueden estar sobreestimados. Estos resultados, por tanto, se oponen con contundencia a la hipótesis de que estudiar en universidades privadas suponga una ventaja sustancial en términos de ingresos.

### *Análisis de sensibilidad de la regresión*

Nunca podremos estar seguros de si un conjunto concreto de controles bastará para eliminar el sesgo de selección. Por eso es importante preguntarse en qué medida los resultados de una regresión son sensibles a cambios en el conjunto de controles. La confianza en las estimaciones por regresión de los efectos causales aumenta cuando los efectos del tratamiento se hacen insensibles (los maestros dicen que son «robustos») sobre el añadido o la eliminación de una variable particular, mientras que algunos controles determinados se mantengan en el modelo. Este patrón deseable queda ilustrado en las columnas (4) a (6) de las tablas 2.2 y 2.3, que presentan estimaciones que apuntan a que la ventaja de estudiar en un centro privado es insensible a la inclusión de las capacidades del alumnado (medidas mediante las notas en las pruebas de acceso a la universidad), ingresos familiares y otras variables de control, una vez que se ha tenido en

cuenta el carácter de los centros en los que el alumnado presentó su solicitud de ingreso.

La fórmula del SVO explica este hallazgo tan llamativo. Comencemos con la tabla 2.5, que muestra los coeficientes de ecuaciones semejantes a la (2.4), salvo en que en lugar de $FS_i$ se introdujo $SAT_i$ en el primer miembro para generar las estimaciones de las columnas (1)-(3), mientras que al situar $\ln P_i$ en el primer miembro resultan las columnas (4)-(6). Estas regresiones auxiliares esclarecen la relación (condicionada a otros controles del modelo) entre cursar estudios en centros privados y dos de nuestros controles, $SAT_i$ y $\ln P_i$. No sorprende observar que acudir a un centro privado es una buena variable predictora tanto de las puntuaciones en los exámenes de acceso a la universidad como de los ingresos familiares, relaciones que quedan documentadas en las columnas (1) y (4) de la tabla. Cuando se añaden controles demográficos, resultados en los estudios de secundaria y una variable binaria que refleja la participación en equipos deportivos, los resultados cambian muy poco, como se ve en las columnas (2) y (5). Pero si se incluyen controles sobre el número de solicitudes cursadas y la nota media en las pruebas de acceso en las universidades solicitadas, como en el modelo de autorrevelado, se elimina de manera muy efectiva la relación entre el hecho de acudir a un centro privado y estas importantes variables de trasfondo. Esto explica por qué los coeficientes estimados para el efecto de estudiar en universidades privadas son esencialmente los mismos en las columnas (4), (5) y (6) de la tabla 2.3.

La fórmula del SVO es la Primera Directiva de la econometría aplicada, así que breguemos con los números para ver cómo entra en acción. A modo de ilustración tomaremos como modelo simple una regresión que introduzca en $P_i$ el logaritmo de los ingresos, sin controles, y como modelo largo, una regresión que tenga en cuenta la calificación individual en las pruebas de acceso a la universidad (variable $SAT_i$). El coeficiente de $P_i$ que se deduce de la regresión simple (sin controles) consta en la columna (1) de la tabla 2.3 y asciende a 0,212, mientras que el coeficiente correspondiente de la regresión larga (que incluye el control $SAT_i$) aparece en la columna (2) y vale 0,152. Como puede verse también en la columna (2) de la tabla, el efecto de $SAT_i$ en la regresión larga asciende a 0,051. La primera columna de la tabla 2.5 muestra que la regresión de la variable omitida

## Tabla 2.5. Efectos de la universidad privada: sesgo de variables omitidas

| | Variable dependiente | | | | | |
|---|---|---|---|---|---|---|
| | Nota de acceso individual ÷ 100 | | | Logaritmo de los ingresos paternos | | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Centro privado | 1.165 | 1.130 | .066 | .128 | .138 | .028 |
| | (.196) | (.188) | (.112) | (.035) | (.017) | (.037) |
| Mujeres | | -.367 | | | .016 | |
| | | (.076) | | | (.013) | |
| Negros | | -1.947 | | | -.359 | |
| | | (.079) | | | (.019) | |
| Hispanos | | -1.185 | | | .259 | |
| | | (.168) | | | (.050) | |
| Asiáticos | | -014 | | | -.060 | |
| | | (.116) | | | (.031) | |
| Otra raza (o no consta) | | -.521 | | | .082 | |
| | | (.293) | | | (.061) | |
| En el 10% mejor de su escuela secundaria | | .948 | | | -.066 | |
| | | (.107) | | | (.011) | |
| No consta nivel de secundaria | | .556 | | | -.030 | |
| | | (.102) | | | (.023) | |
| Deportista | | -.318 | | | .017 | |
| | | (.147) | | | (.016) | |
| Puntuación media en pruebas de acceso en las escuelas solicitadas ÷ 100 | | | .777 | | | .063 |
| | | | (.058) | | | (.014) |
| Presentaron dos solicitudes | | | .252 | | | .020 |
| | | | (.077) | | | (.010) |
| Presentaron tres solicitudes | | | .375 | | | .042 |
| | | | (.106) | | | (.013) |
| Presentaron cuatro o más solicitudes | | | .330 | | | .079 |
| | | | (.093) | | | (.014) |

*Notas:* Esta tabla describe la relación entre el hecho de cursar estudios en una universidad privada y las características personales. Las variables dependientes son la nota obtenida en las pruebas de acceso a la universidad (dividida entre 100) en las columnas (1)-(3) y el logaritmo de los ingresos de los padres en las columnas (4)-(6). Cada columna muestra el coeficiente de una regresión de la variable dependiente sobre una variable binaria que señala si se asistió o no a un centro privado, así como varios controles. El tamaño de la muestra es 14.238. Los errores típicos constan entre paréntesis.

$SAT_i$ sobre la incluida $P_i$ conduce a un coeficiente de 1,165. Si se reúne esta información se puede deducir el SVO de dos maneras:

$$\begin{aligned} SVO &= Simple - Larga = 0{,}212 - 0{,}152 = 0{,}06 \\ SVO &= [Regresión\ de\ omitida\ sobre\ incluida] \times \\ &\quad \times [Efecto\ de\ la\ omitida\ en\ larga] = \\ &= 1{,}165 \times 0{,}051 = 0{,}06. \end{aligned}$$

Compárese lo anterior con el cálculo paralelo que nos lleva desde la columna (4) hasta la (5) en la tabla 2.3. Estas columnas reflejan los resultados de modelos que incluyen controles de autorrevelado. Aquí *Simple-Larg*a tiene un valor reducido, 0,034 – 0,031 = 0,003 para ser exactos. Tanto la regresión simple como la larga incluyen controles selectivos procedentes del modelo de autorrevelado, al igual que la relevante regresión auxiliar de la puntuación individual en las pruebas de acceso a la universidad, $SAT_i$, sobre $P_i$. Cuando se incluyen controles de autorrevelado en ambos modelos se deduce:

$$\begin{aligned} SVO &= [Regresión\ de\ omitida\ sobre\ incluida] \times \\ &\quad \times [Efecto\ de\ la\ omitida\ en\ larga] = \\ &= 0{,}066 \times 0{,}036 = 0{,}0024. \end{aligned}$$

(El error de redondeo en números pequeños nos desvía del objetivo de 0,003). El efecto de omitir $SAT_i$ en la regresión larga cae ahora de 0,051 a 0,036, mientras que la regresión de la variable omitida sobre la incluida cae en un orden de magnitud, desde el abultado valor 1,165 hasta 0,066 (consta en la columna [3] de la tabla 2.5). Esto indica que, si se tienen en cuenta tanto el número de escuelas solicitadas como la selectividad media de las mismas, entonces no se aprecian diferencias significativas entre los estudiantes que eligen centros públicos y los que optan por los privados, al menos en lo que respecta a las calificaciones obtenidas en las pruebas de acceso a la universidad. En consecuencia, desaparece el contraste entre la estimación simple y la larga.

El efecto de estudiar en un centro privado resulta insensible a la inclusión de las variables disponibles sobre capacidades individuales y sobre trasfondo familiar, una vez que se incluyen controles de

autorrevelado. Del mismo modo, otras variables de control, incluyendo aquellas sobre las que no disponemos de datos, podrían tener también muy poca influencia. En otras palabras, es probable que cualquier SVO debido a diferencias no controladas sea modesto.[14] Este indicio circunstancial acerca de la escasa importancia del SVO no garantiza que los resultados de las regresiones tratadas en este capítulo posean la misma fuerza causal que los que se obtendrían de un experimento aleatorio: siempre preferiríamos disponer de un experimento real. Sin embargo, y como mínimo, estos resultados ponen en cuestión la afirmación de que estudiar en las caras universidades privadas confiere una ventaja sustancial en cuanto a los ingresos futuros.

Maestro Stevefu: Por favor, Pequeño Saltamontes, sé breve.

Pequeño saltamontes: Las comparaciones causales comparan lo semejante con lo semejante. Para esclarecer los efectos de la elección de universidad nos centramos en estudiantes de características similares.

Maestro Stevefu: Cada cual puede ser diferente de un millar de maneras. ¿Han de ser similares en todo?

Pequeño Saltamontes: Las comparaciones buenas eliminan las diferencias sistemáticas entre quienes eligieron un camino y quienes optaron por otro, si tales diferencias están asociadas a los ingresos.

Maestro Stevefu: ¿Cómo puede lograrse eso?

Pequeño Saltamontes: El método de emparejamiento reúne individuos en grupos concordes, con los mismos valores de las variables de control, como medidas de las capacidades individuales o del entorno familiar. Las comparaciones concordes dentro de esos grupos se promedian después para deducir un efecto global único.

[14] Joseph Altonji, Todd Elder y Christopher Taber formalizan la noción de que el SVO asociado a los regresores accesibles proporciona una guía acerca del SVO generado por los regresores de los que no se dispone. Véanse los detalles en su estudio «Selection on Observed and Unobserved Variables: Assesing the Efectiveness of Catholic Schools», *Journal of Political Economy*, vol. 113, número 1, febrero de 2005, páginas 151-184.

Maestro Stevefu: ¿Y la regresión?

Pequeño Saltamontes: La regresión empareja de manera automática. La estimación por regresión de un efecto causal constituye también un promedio de las comparaciones dentro de cada grupo.

Maestro Stevefu: ¿Cuál es el tao del SVO?

Pequeño Saltamontes: El SVO es la diferencia entre los coeficientes de la regresión simple y la larga. La regresión larga incluye controles adicionales que se omiten en la simple. La simple es igual a la larga más el efecto en la larga de la variable omitida multiplicado por la regresión de la omitida sobre la incluida.

Maestro Joshway: Aquí no se ha omitido nada, Pequeño Saltamontes.

## Maestros de la econometría: Galton y Yule

El término *regresión* lo acuñó sir Francis Galton, medio primo de Charles Darwin, en 1886. Galton era un hombre de amplios intereses, pero quedó muy impactado por la obra maestra de Darwin, *El origen de las especies*. Galton confiaba en aplicar la teoría evolutiva de Darwin a la variación de los rasgos humanos. En el curso de sus investigaciones estudió atributos que abarcan desde las huellas dactilares a la belleza. Fue además uno de los muchos intelectuales británicos que pusieron la obra de Darwin al servicio de la siniestra eugenesia. A pesar de esta lamentable desviación, su trabajo en estadística teórica ejerció efectos duraderos y saludables sobre las ciencias sociales. Galton estableció los cimientos estadísticos de las ciencias sociales cuantitativas como las que ahora nos ocupan.

Galton descubrió que las estaturas medias de padres e hijos están ligadas a través de una ecuación de regresión. También desveló una consecuencia interesante de este modelo particular de regresión: la estatura media de los hijos es una media ponderada de la estatura de los padres y de la estatura media de la población de la que proceden los padres y los hijos. En consecuencia, los progenitores más altos que la media tendrán descendientes no tan altos, mientras que los progenitores más bajos que la media tendrán descendientes no tan bajos. En concreto, el maestro Stevefu, que mide 191 centímetros, puede esperar que su progenie tenga una buena estatura, pero no tanto como la suya. En cambio, y por fortuna, el maestro Joshway, que en días fa-

vorables mide 168 centímetros, puede esperar que sus descendientes alcancen una estatura algo mayor.

Galton explicó este fenómeno de promediado en su célebre artículo de 1886 titulado «Regression towards Mediocrity in Hereditary Stature».[15] Hoy día esta propiedad recibe el nombre de *regresión a la media*. La regresión a la media no constituye una relación causal, sino que se trata de una propiedad estadística de los pares de variables correlacionadas, tales como las estaturas de padres e hijos. Aunque las estaturas de padres e hijos nunca sean exactamente iguales, sus distribuciones de frecuencias se mantienen esencialmente invariables. Esta estabilidad de las distribuciones es la causa de la regresión de Galton.

Vemos en la regresión un procedimiento estadístico que nos confiere el poder de efectuar comparaciones más igualadas, mediante la introducción de variables de control en los modelos, para esclarecer los efectos de un tratamiento. Galton no parecía tener interés en la regresión como estrategia de control. El iniciador del uso de la regresión con fines de control fue George Udny Yule, alumno del estadístico Karl Pearson, un protegido de Galton. Yule se dio cuenta de que el método de regresión de Galton podría extenderse para incorporar muchas variables. En un artículo de 1899, Yule usó esta extensión para estudiar la relación entre la aplicación de las leyes inglesas de pobreza *(Poor Laws)* en los distintos condados, y la probabilidad de que los residentes de esos condados fueran pobres, introduciendo controles sobre el crecimiento de la población y la distribución de edades en cada condado.[16] Las leyes de pobreza garantizaban la subsistencia de los indigentes, normalmente proporcionándoles alojamiento y empleo en unos albergues denominados *workhouses*.[17] Yule tenía un interés especial en averiguar si la práctica de la llamada *atención externa*, que implicaba proporcionar ayudas en metálico pero sin exigir la entrada en un albergue, aumentaba los índices de pobreza al hacer más llevadera la condición de pobre. He aquí una cuestión causal

[15] Francis Galton, «Regression towards Mediocrity in Hereditary Stature», *Journal of the Anthropological Institute of Great Britain and Ireland*, vol. 15, 1886, páginas 246-263.

[16] George Udny Yule, «An Investigation into the Causes of Changes in Pauperism in England, Chiefly during the Last Two Intercensal Decades», *Journal of the Royal Statistical Society*, vol. 62, número 2, enero de 1899, páginas 249-295.

[17] Literalmente, «casas de trabajo». *(N. de la T.)*

bien planteada en unos términos semejantes a aquellas de las que se ocupa la ciencia social de hoy.

## Apéndice: Teoría de la regresión

### *Funciones de valor esperado condicionado*

En el capítulo 1 se introdujo el concepto de *valor esperado*, a veces llamado *esperanza* para abreviar. Escribimos $E[Y_i]$ para referirnos al valor esperado de la variable $Y_i$. También nos interesa el *valor esperado condicionado*, es decir, el valor esperado de una variable dentro de un grupo (también llamado *celda*) definido por una segunda variable. A veces, esta segunda variable es de carácter binario y adopta sólo dos posibles valores, pero no siempre tiene por qué ser así. Con frecuencia, como sucede en este capítulo, nos interesan los valores esperados condicionados para grupos definidos según los valores de variables que no son binarias como, por ejemplo, los ingresos esperados de personas que han completado 16 años de formación académica. Este tipo de valor esperado condicionado se puede escribir como

$$E[Y_i|X_i = x],$$

y se lee como «el valor esperado de $Y_i$, cuando $X_i$ adopta el valor concreto $x$».

Los valores esperados condicionados nos dicen cómo cambia la media poblacional de una variable a medida que alteramos la variable condicional dentro del abanico de valores que puede adoptar. Para cada valor de la variable condicional podemos obtener un promedio distinto de la variable dependiente, $Y_i$. El conjunto de tales promedios se denomina la *función de valor esperado condicionado* (FVEC abreviado); $E[Y_i|X_i]$ es la FVEC de $Y_i$ conocido $X_i$, pero sin especificar un valor para $X_i$, mientras que $E[Y_i|X_i = x]$ especifica un punto concreto en el dominio de esta función.

La figura 2.1 muestra una de nuestras FVEC favoritas. Los puntos de la gráfica representan el promedio del logaritmo de los ingresos mensuales para hombres con niveles de escolaridad diferentes (medidos según el último curso superado), los cuales se describen en el eje

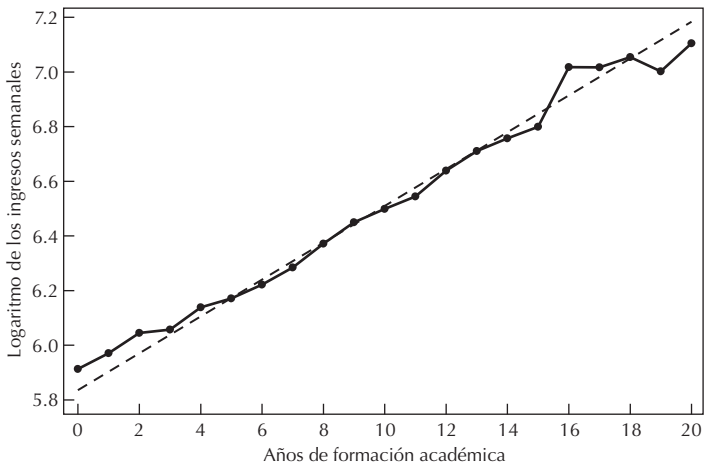



**Figura 2.1.** La FVEC y la recta de regresión.
*Notas:* Esta figura muestra la función de valor esperado condicionado (FVEC) del logaritmo de los ingresos semanales según los años de formación, así como la línea recta generada al efectuar una regresión del logaritmo de los ingresos semanales sobre los años de formación (representada con trazo discontinuo).

$X$ (los datos proceden del censo de Estados Unidos de 1980). Aunque presenta ciertas irregularidades, la FVEC de ingresos sobre escolaridad muestra una tendencia ascendente muy marcada, con una pendiente promedio de alrededor de 0,1. Dicho de otro modo, cada año de escolaridad va asociado a ingresos que son un 10% superiores en promedio.

Muchas de las FVEC que nos interesan involucran más de una variable condicional, cada una de las cuales puede adoptar dos o más valores. Escribimos

$$E[Y_i|X_{1i},\ldots,X_{Ki}]$$

para una FVEC con $K$ variables condicionales. Cuesta más representar gráficamente una FVEC con muchas variables condicionales, pero la idea es la misma. $E[Y_i|X_{1i},\ldots,X_{Ki}]$ da el promedio poblacional de $Y_i$ cuando estas otras $K$ variables se mantienen fijas. En lugar de fijarnos en los ingresos promedio condicionados sólo a la escolaridad, podríamos considerar también celdas condicionales definidas, por ejemplo, por edad, raza o género.

### *Regresión y la FVEC*

La tabla 2.1 ilustra la filosofía del emparejamiento al comparar estudiantes que asistieron a centros públicos o privados, habiendo distribuido previamente a los estudiantes en celdas según las universidades en las que presentaron sus solicitudes, y en las que fueron admitidos. El grueso del capítulo se dedica a explicar cómo la regresión ofrece un modo rápido y fácil de automatizar esas comparaciones en grupos de individuos concordes. Aquí usaremos la FVEC para hacer una interpretación más rigurosa de la regresión.[18]

Las estimaciones por regresión de la ecuación (2.2) que constan en la tabla 2.3 sugieren que acudir a una universidad privada carece de relación con los ingresos medios, una vez se han fijado la nota individual en las pruebas de acceso, los ingresos paternos y el nivel selectivo de las universidades que se solicitan y en las que se ingresa. Como simplificación supongamos que la FVEC del logaritmo de los ingresos fuera una función lineal de estas variables condicionales. En concreto admitamos que

$$E[\text{en } Y_i | P_i, GROPUP_i, SAT_i, \text{en } PI_i] = \tag{2.6}$$
$$= \alpha + \beta Pi + \sum_j \gamma_j GROPUP_{ji} + \delta_1 SAT_i + \delta_2 \text{ en } PI_i,$$

donde las letras griegas, como siempre, son parámetros. Si la FVEC de $\ln Y_i$ es una función lineal de las variables condicionales, como en la ecuación (2.6), la regresión de $\ln Y_i$ sobre esas mismas variables condicionales estima esa función lineal. (No damos una prueba detallada de este hecho, aunque no es complicada.) En particular, si hay linealidad, el coeficiente de $P_i$ en la ecuación (2.2) será igual al coeficiente de $P_i$ en la ecuación (2.6).

Con una FVEC lineal, las estimaciones por regresión de los efectos de estudiar en un centro privado basados en la ecuación (2.2) también resultan idénticas a las que se obtendrían mediante una estrategia que (i) agrupara a los estudiantes según los valores de $GROUP_i$, $SAT_i$ y $\ln P_i$; (ii) comparara dentro de cada grupo los ingresos promedio de los estudiantes que asistieron a centros privados ($P_i = 1$) con

[18] Puede consultarse una explicación más detallada en el capítulo 3 de Angrist y Pischke, Mostly Harmless Econometrics, 2009.

los que acudieron a centros públicos ($P_i = 0$) para cada combinación posible de las variables condicionales, y (iii) produjera un promedio simple promediando todas estas comparaciones concretas dentro de las celdas. Para ver esto basta usar la ecuación (2.6) y escribir las comparaciones específicas dentro de cada celda como

$$E[\text{en } Y_i|P_i = 1,\ GROPUP_i,\ SAT_i, \text{ en } PI_i] =$$
$$-E[\text{en } Y_i|P_i, = 0,\ GROPUP_i,\ SAT_i, \text{ en } PI_i] = \beta.$$

Como nuestro modelo lineal para la FVEC admite que el efecto de acudir a un centro privado es igual a la constante $\beta$ en todas las celdas, cualquier promedio ponderado de comparaciones concretas de celda sobre este aspecto también resultará igual a $\beta$.

Los modelos lineales nos ayudan a comprender la regresión, pero la regresión es una herramienta fabulosamente flexible, que puede emplearse con independencia de que la FVEC subyacente sea o no lineal. La regresión hereda esta flexibilidad del siguiente par de propiedades teóricas muy relacionadas entre sí:

- Si $E[Y_i|X_{1i},\ldots,X_{Ki}] = a + \sum_{k=1}^{K} b_k X_{ki}$ para algunas constantes $a$ y $b_1,\ldots,b_K$, entonces la regresión de $Y_i$ sobre $X_{1i},\ldots,X_{Ki}$ tiene ordenada en el origen $a$ y pendientes $b_1,\ldots,b_K$. Es decir, si la FVEC de $Y_i$ condicionada a $X_{1i},\ldots,X_{Ki}$ es lineal, entonces la regresión de $Y_i$ sobre $X_{1i},\ldots,X_{Ki}$ nos estima esa función.
- Si $E[Y_i|X_{1i},\ldots,X_{Ki}]$ es una función no lineal de las variables condicionales, entonces la regresión de $Y_i$ sobre $X_{1i},\ldots,X_{Ki}$ constituye la mejor aproximación lineal a esta FVEC no lineal en el sentido de que minimiza la desviación cuadrática media esperada entre los valores ajustados mediante un modelo lineal y la FVEC.

En resumen: si la FVEC es lineal, la regresión da con ella; si no es lineal, la regresión encuentra una buena aproximación a la misma. Acabamos de usar la primera propiedad teórica para interpretar las estimaciones por regresión de los efectos de las universidades privadas cuando la FVEC es lineal. La segunda propiedad nos dice que podemos esperar que las estimaciones mediante una regresión de

los efectos de un tratamiento estén cerca de las que obtendríamos agrupando de acuerdo con los valores de las variables y luego promediando las diferencias entre el tratamiento y el control de cada celda, incluso aunque la FVEC no sea lineal.

La figura 2.1 ilustra el modo en que la regresión se aproxima a la FVEC no lineal del logaritmo de los ingresos condicionada a la formación académica. Aunque la FVEC oscile en torno a la línea de regresión, esta línea capta la fuerte relación positiva que existe entre la formación académica y los salarios. Además, la pendiente de la regresión resulta cercana a $E\{E[Y_i|X_i] - E[Y_i|X_i - 1]\}$, es decir, la pendiente de la regresión también se acerca al efecto esperado de un cambio del valor de $X_i$ en una unidad en $E[Y_i|X_i]$.[19]

### *Regresión simple y covarianza*

La regresión está estrechamente ligada al concepto estadístico de *covarianza* (o *covariancia*). La covarianza entre dos variables, $X_i$ y $Y_i$, se define como

$$C(X_i,Y_i) = E\Big[\big(X_i - E[X_i]\big)\big(Y_i - E[Y_i]\big)\Big].$$

La covarianza tiene tres propiedades importantes:

(i) La covarianza de una variable consigo misma es su varianza; $C(X_i,X_i) = \sigma_X^2$.
(ii) Si el valor esperado de una de las dos variables, $X_i$ o $Y_i$, es 0, entonces la covarianza entre ellas es el valor esperado de su producto; $C(X_i,Y_i) = E[X_iY_i]$.
(iii) La covarianza entre funciones lineales de las variables $X_i$ e $Y_i$ (definidas como $W_i = a + bX_i$, $Z_i = c + dY_i$) mediante las constantes $a$, $b$, $c$, $d$, viene dada por

$$C(W_i, Z_i) = bd\, C(X_i,Y_i).$$

[19] La cantidad entre corchetes, $E[Y_i|X_i] - E[Y_i|X_i - 1]$, es una función de $X_i$ y, por tanto, al igual que la variable $X_i$, posee un valor esperado.

La conexión íntima entre la regresión y la covarianza se puede ver en un *modelo de regresión simple*, es decir, una regresión con un regresor, $X_i$, más una ordenada en el origen.[20] La pendiente y la ordenada en el origen de la regresión simple son los valores de $a$ y $b$ que minimizan la correspondiente *suma cuadrática de residuos* (SCR), que escribimos como

$$SCR(a,b) = E[(Y_i - a - bX_i)^2].$$

La expresión SCR se refiere a la suma de cuadrados, porque a la hora de efectuar la minimización en una muestra concreta se sustituye el valor esperado por una media simple, o una suma. La solución en el caso de dos variables es

$$b = \beta = \frac{C(Y_i, X_i)}{V(X_i)} \tag{2.7}$$

$$\alpha = \alpha = E[(Y_i] - \beta E[Y_i].$$

La ecuación (2.7) implica que cuando hay dos variables *no correlacionadas* (con covarianza mutua 0), la regresión de cualquiera de ellas sobre la otra arroja una pendiente nula. Del mismo modo, una pendiente en una regresión simple nula significa que las dos variables no están correlacionadas.

### *Ajustes y residuos*

La regresión separa cualquier variable dependiente en dos partes. En concreto, para la variable dependiente $Y_i$ podemos escribir

$$Y_i = \hat{Y}_i + e_i.$$

[20] Mientras que en inglés, cuando en la regresión intervienen dos variables, una dependiente en el primer miembro y una explicativa en el segundo miembro, se suele hablar, en traducción literal, de regresión «bivariada» hemos preferido mantener el uso habitual en castellano y hablar de regresión «simple» A los modelos de regresión «multivariada», que añaden más variables explicativas a este esquema básico, los llamamos de regresión «múltiple», siguiendo la costumbre en castellano. *(N. del E.)*

El primer término, $\hat{Y}_i$, representa los valores ajustados y se acostumbra a decir que es la parte de $Y_i$ «explicada» por el modelo. La segunda parte, $e_i$, el residuo, es lo que sobra.

Los residuos de la regresión y los regresores incluidos en el modelo que los produce son cantidades no correlacionadas. Es decir, si $e_i$ es el residuo de una regresión con variables $X_{1i},\ldots,X_{Ki}$, entonces la regresión de $e_i$ con estas mismas variables producirá coeficientes que serán todos ellos nulos. Como los valores estimados son una combinación lineal de los regresores, se concluye que tampoco están correlacionados con los residuos. Pasemos a resumir estas importantes propiedades.

Propiedades de los residuos Supongamos que $\alpha$ y $\beta_1,\ldots,\beta_K$ son la ordenada en el origen y las pendientes que resultan de efectuar la regresión de $Y_i$ sobre $X_{1i},\ldots,X_{Ki}$. Los valores estimados de esta regresión son:

$$\hat{Y}i = \alpha + \sum_{k=1}^{K} \beta_k X_{ki},$$

y los correspondientes residuos de la regresión son:

$$e_i = Y_i - \hat{Y}_i = Y_i - \alpha - \sum_{k=1}^{K} \beta_k X_{ki}.$$

Los residuos de la regresión

(i) tienen valor esperado y media muestral nulos: {insertar aquí la fórmula con la que termina el punto (i), justo al final de la página 87};

(ii) no están correlacionados, tanto a nivel de población como de muestra, con los regresores de los que se derivan, ni con los valores estimados correspondientes. Es decir, para cada regresor $X_{ki}$,

$$E[X_{ki}e_i] = \sum_{i=1}^{n} X_{ki}e_i = 0;\ E[\hat{Y}_i e_i] = \sum_{i=1}^{n} \hat{Y}_i e_i = 0,$$

Estas propiedades podrían parecernos un artículo de fe, pero son fáciles de deducir si se sabe un poco de análisis matemático. Se parte

del hecho de que los parámetros de la regresión y los valores estimados minimizan la suma de residuos cuadráticos. Las condiciones de primer orden de este problema de minimización son equivalentes a los puntos (i) y (ii) anteriores.

### *Regresión con variables binarias*

Un caso especial y muy importante de regresión es el de la regresión simple en la que uno de los regresores es una variable binaria. El valor esperado condicionado de $Y_i$ dado el valor de una variable binaria, $Z_i$, puede adoptar dos valores. Escritos con letras griegas se verían así:

$$\begin{aligned} E[Y_i|Z_i = 0] &= \alpha \\ E[Y_i|Z_i = 1] &= \alpha + \beta, \end{aligned}$$

de manera que

$$\beta = E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]$$

representa el cambio del valor esperado de $Y_i$ cuando el regresor binario $Z_i$ se activa o se desactiva.

Con esta notación podemos escribir

$$\begin{aligned} E[Y_i|Z_i] &= E[Y_i|Z_i = 0] + \big(E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]\big)Z_i \\ &= \alpha + \beta Z_i. \end{aligned} \tag{2.8}$$

Esto indica que $E[Y_i|Z_i]$ es una función lineal de $Z_i$, con pendiente $\beta$ y ordenada en el origen $\alpha$. Como la FVEC con una única variable binaria es lineal, la regresión ajusta esta función perfectamente. En consecuencia, la pendiente de la regresión debe ser también $\beta = E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]$, la diferencia entre los valores esperados de $Y_i$ cuando $Z_i$ se activa o desactiva.

La importancia de la regresión con variables binarias se debe a la frecuencia con que encontramos este tipo de regresores, como en nuestros análisis de los seguros médicos o de los tipos de centros educativos en los que se cursan los estudios.

### *Anatomía de la regresión y la fórmula del SVO*

Las regresiones más interesantes son las múltiples, es decir, las que incluyen una variable causal de interés y una o más variables de control. Por ejemplo, la ecuación (2.2) plantea una regresión del logaritmo de los ingresos sobre una variable binaria que marca si se ha asistido o no a una universidad privada, en un modelo que incorpora controles para las aptitudes individuales, el contexto familiar y el nivel selectivo de los centros en los que se solicita la admisión y en los que se logra el acceso. Ya hemos explicado que el control de las variables explicativas se parece al proceso de emparejamiento. Es decir, el coeficiente de regresión de una variable binaria que indique el acceso a una universidad privada en un modelo que incluya controles es similar al que se obtendría si distribuyéramos a los estudiantes en celdas sobre la base de esos mismos controles, comparáramos a los estudiantes de centros públicos y privados dentro de esas celdas, y luego tomáramos el promedio del conjunto resultante de las comparaciones condicionadas. A continuación, ofrecemos una lección de «anatomía de la regresión» más detallada.

Supongamos que la variable causal de interés es $X_{1i}$ (podría ser una variable binaria que indicase si se estudió o no en un centro privado), y que la variable de control es $X_{2i}$ (podría ser la nota en las pruebas de acceso a la universidad). Tras cierto esfuerzo se puede llegar a la expresión siguiente para el coeficiente de $X_{1i}$ en una regresión con $X_{2i}$ como control:

$$\beta_1 = \frac{C(Y_i, \widetilde{X}_{1i})}{V(\widetilde{X}u)},$$

donde $\widetilde{X}_{1i}$ representa el residuo de la regresión de $X_{1i}$ sobre $X_{2i}$:

$$X_{1i} = \pi_0 + \pi_1 X_{2i} + \widetilde{X}_{1i}.$$

Como siempre, los residuos no están correlacionados con los regresores de los que proceden, y lo mismo ocurre para el residuo $\widetilde{X}_{1i}$. No deber sorprender, por tanto, que el coeficiente de $X_{1i}$ en una regresión múltiple que incluya $X_{2i}$ como control sea el coeficiente bivariado procedente de un modelo que incluya sólo la parte de $X_{1i}$ que no esté correlacionada con $X_{2i}$. Esta importante ecuación de la anatomía de

la regresión determina nuestra interpretación de los coeficientes de una regresión.

La idea de la anatomía de la regresión se extiende a modelos con más de dos regresores. El coeficiente multivariado de un regresor dado se puede escribir como el coeficiente de una regresión simple del residuo de ese regresor sobre al resto de regresores. Veamos la anatomía del *k*-ésimo coeficiente en un modelo con *K* regresores:

### Anatomía de la regresión

$$\beta_k = \frac{C(Y_i, \tilde{X}_{ki})}{V(\tilde{X}_{ki})},$$

donde $\tilde{X}_{ki}$ es el residuo de una regresión de $X_{ki}$ sobre las otras $K - 1$ variables explicativas incluidas en el modelo.

La anatomía de la regresión resulta especialmente reveladora cuando los controles son variables binarias, como en la ecuación (2.2). Para los fines de esta explicación podemos simplificar el modelo de interés para que queden sólo variables de control binarias, es decir,

$$\text{en } Y_i = \alpha + \beta P_i + \sum_{j=1}^{150} \gamma_j GROUP_{ji} + e_j. \tag{2.9}$$

La anatomía de la regresión nos dice que el coeficiente de $P_i$ en el modelo con 150 variables binarias de control $GROUP_{ji}$ es el coeficiente que resultaría de una regresión de $\tilde{P}_i$, símbolo que representa el residuo de una regresión de $P_i$ sobre el conjunto constante de las 150 variables binarias $GROUP_{ji}$.

Será de ayuda añadir ahora un segundo subíndice para señalar los grupos, y no sólo los individuos. En este esquema, $\ln Y_{ij}$ es el logaritmo de los ingresos del graduado universitario $i$ en el grupo de selectividad $j$, mientras que $P_{ij}$ representa el tipo de escolarización universitaria (privada o pública) de este mismo graduado. ¿Cuál es el residuo $\tilde{P}_{ij}$ que resulta de la regresión auxiliar de $P_{ij}$ sobre el conjunto de 150 variables binarias de grupos selectivos? Como la regresión auxiliar que genera $\tilde{P}_{ij}$ incluye un parámetro para cada posible valor de la FVEC subyacente, esta regresión reproduce a la perfección la FVEC de $P_{ij}$ condicionada al grupo selectivo. (Aquí estamos exten-

diendo el resultado acerca de las variables binarias descrito por la ecuación [2.8] a regresiones sobre variables binarias que describen una variable categorizada que adopta muchos valores, y no sólo dos.) En consecuencia, el valor estimado por la regresión de $P_{ij}$ sobre todo el conjunto de variables binarias de grupo selectivo será el nivel medio de asistencia a escuelas privadas dentro de cada grupo. Para el solicitante *i* en el grupo *j*, el residuo de la regresión auxiliar será por tanto $\tilde{P}_{ij} = P_{ij} - \bar{P}_j$, donde $\bar{P}_j$ representa la media de asistencia a centros privados dentro del grupo selectivo *j*, al que pertenece el sujeto *i*.

Para terminar, si se reúnen todas las piezas, la anatomía de la regresión nos dice que el parámetro $\beta$ en el modelo de regresión múltiple descrito por la ecuación (2.9) es:

$$\beta = \frac{C(\text{en } Y_{ij}, \tilde{P}_{ij})}{V(\tilde{P}_{ij})} = \frac{C(\text{en } Y_{ij}, P_{ij} - \bar{P}_{ij})}{V(P_{ij} - \bar{P}_{ij})}. \qquad (2.10)$$

Esta expresión pone de manifiesto que la regresión sobre la asistencia a universidades privadas con controles binarios para grupos selectivos es también un proceso que se realiza dentro de cada grupo, justo igual que si hubiéramos organizado a mano a los estudiantes en grupos y hubiéramos comparado a los estudiantes de centros públicos y privados dentro de cada grupo: la variación entre grupos se elimina al restar $\bar{P}_j$ para construir los residuos $\tilde{P}_{ij}$. Además, como sucede con los grupos C y D de la tabla 2.1, la ecuación (2.10) implica que los grupos de solicitantes en los que todos sus integrantes acuden, o bien a un centro público, o bien a uno privado, no aportan información sobre los efectos de estudiar en universidades privadas, porque $P_{ij} - \bar{P}_j$ es cero para todos los miembros de tales grupos.

La fórmula del SVO, usada al final de este capítulo (apartado 2.3) para interpretar las estimaciones de modelos con conjuntos diferentes de controles, ilustra otra prestación reveladora de la anatomía de la regresión. Llamemos $\beta^l$, coeficiente de regresión larga, al coeficiente de $X_{1i}$ en un modelo de regresión múltiple que incluya $X_{2i}$ como control:

$$Y_i = \alpha^l + \beta^l X_{1i} + \gamma X_{2i} + e_i^l.$$

Y llamemos $\beta^s$ al coeficiente de $X_{1i}$ en un modelo de regresión simple (es decir, que no incluye $X_{2i}$):

$$Y_i = \alpha^s + \beta^s X_{1i} + e_i^l.$$

La fórmula del SVO describe la relación entre los coeficientes de las regresiones simple y larga del modo siguiente.

fórmula del sesgo de variables omitidas (svo)

$$\beta^s = \beta^l + \pi_{21}\gamma,$$

donde $\gamma$ es el coeficiente de $X_{2i}$ en la regresión larga, y $\pi_{21}$ el coeficiente de $X_{1i}$ en una regresión de $X_{2i}$ sobre $X_{1i}$.[21] Si se expresa con palabras: *simple igual a larga más el efecto de la omitida por la regresión de la omitida sobre la incluida.*

Vale la pena deducir esta fórmula crucial. La pendiente del modelo simple es:

$$\beta^s = \frac{C(Y_i, X_{1i})}{V(X_{1i})}. \tag{2.11}$$

Si en lugar de $Y_i$ se coloca el modelo largo en la ecuación (2.11), se obtienen

$$\begin{aligned}
&\frac{C(\alpha^l + \beta_1^l X_{1i} + \gamma X_{2i} + e_i^l, X_{1i})}{V(X_{1i})} \\
&= \frac{\beta_1^l V(X_{1i}) + \gamma C(X_{2i}, X_{1i}) + (e_i^l, X_{1i})}{V(X_{1i})} \\
&= \beta_1^l + \frac{C(X_{2i}, X_{1i})}{V(X_{1i})}\gamma = \beta_1^l + \pi_{21}\gamma.
\end{aligned}$$

La primera igualdad procede del hecho de que la covarianza de una combinación lineal de variables es la correspondiente combinación lineal de covarianzas, tras recolocar los términos. También, la covarianza de una constante con cualquier otra cosa es cero, y la covarianza de una variable consigo misma es la varianza de la variable. La segunda igualdad se justifica porque $C(e_i^l, X_{1i}) = 0$, dado que los resi-

[21] La fórmula de la anatomía de la regresión se deduce de un modo similar, así que aquí mostramos todos los pasos solamente para el SVO.

duos no están correlacionados con los regresores de los que proceden ($e_i^l$ es el residuo de una regresión que incluye $X_{1i}$). La tercera igualdad define $\pi_{21}$ como el coeficiente de $X_{1i}$ en una regresión de $X_{2i}$ sobre $X_{1i}$.

Con frecuencia, como en la explicación de las ecuaciones (2.2) y (2.5), nos interesa comparar modelos simples y largos, pero con ciertos controles incluidos en ambos. La fórmula del SVO en este escenario es una extensión trivial de la anterior. Llamemos $\beta^l$, coeficiente de la regresión larga, al coeficiente de $X_{1i}$ en una regresión múltiple que incluya $X_{2i}$ y $X_{3i}$ como controles; denominemos $\beta^s$, coeficiente de la regresión simple, al coeficiente de $X_{1i}$ en una regresión múltiple que incluya sólo $X_{3i}$ como control (y que excluya $X_{2i}$). La fórmula del SVO en este caso se puede seguir escribiendo como

$$\beta^s = \beta^l + \pi_{21}\gamma, \tag{2.12}$$

donde $\gamma$ es el coeficiente de $X_{2i}$ en la regresión larga, pero esa regresión incluye tanto $X_{3i}$ como $X_{2i}$, y $\pi_{21}$ es el coeficiente de $X_{1i}$ en una regresión de $X_{2i}$ sobre $X_{1i}$ y $X_{3i}$ a la vez. De nuevo podemos decir: *corta igual a larga más el efecto de la omitida por la regresión de la omitida sobre la incluida*. Dejamos para el lector la deducción de la fórmula (2.12), un ejercicio que pondrá a prueba su entendimiento de la materia (y constituye una pregunta de examen magnífica).

### *Modelos logarítmicos*

Las regresiones tratadas en este capítulo tienen este aspecto:

$$\text{en } Y_i = \alpha + \beta\, P_i + \sum_j \gamma_j GROUP_{ji} + \delta_1 SAT_i + \delta_2 \text{ en } PI_i + e_i,$$

que repite la ecuación (2.2). ¿Qué hace ahí ese $\ln Y_i$ en el primer miembro? ¿Por qué se introduce el logaritmo, y no la variable original tal cual? La respuesta se comprende mejor si se plantea una regresión simple como

$$\text{en } Y_i = \alpha + \beta\, P_i + e_i, \tag{2.13}$$

donde $P_i$ es una variable binaria que indica si se estudió o no en una universidad privada. Como se trata de un caso de regresión con variables binarias, tenemos que

$$E[\text{en } Y_i | P_i] = \alpha + \beta P_i.$$

Dicho de otro modo, la regresión en este caso estima perfectamente la FVEC.

Supongamos que introducimos una modificación *ceteris paribus* en $P_i$ para el estudiante $i$. Esto conduciría al resultado potencial $Y_{0i}$ cuando $P_i = 0$, y al resultado potencial $Y_{1i}$ si $P_i = 1$. Si volvemos a tomar la ecuación (2.13) como modelo, tendremos para los logaritmos de esos resultados potenciales

$$\begin{aligned} \text{en } Y_{0i} &= \alpha + ei \\ \text{en } Y_{1i} &= \alpha + \beta + e_i. \end{aligned}$$

La diferencia de resultados potenciales es, por tanto,

$$\text{en } Y_{1i} - \text{en } Y_{0i} = \beta. \tag{2.14}$$

Si se reordenan los términos resulta

$$\begin{aligned} \beta = \text{en}\frac{Y_{1i}}{Y_{0i}} &= \text{en}\left(1 + \frac{Y_{1i} - Y_{0i}}{Y_{0i}}\right) \\ &= \text{en}(1 + \Delta\%Y_p) \\ &\approx \Delta\%Y_p, \end{aligned}$$

donde $\Delta\%Y_p$ representa el porcentaje de cambio en el resultado potencial inducido por $P_i$. El análisis matemático nos dice que $\text{en}(1 + \Delta\%Y_p)$ está muy cerca de $\Delta\%Y_p$, si esta última cantidad es pequeña. De aquí se concluye que la pendiente de la regresión en un modelo con en $Y_i$ en el primer miembro da aproximadamente el porcentaje de cambio en $Y_i$ inducido por un cambio en el regresor correspondiente.

Para calcular el porcentaje exacto de cambio inducido por una alteración en $P_i$ tomamos la exponencial de ambos miembros de la ecuación (2.14)

$$\frac{Y_{1i}}{Y_{0i}} = \exp(\beta),$$

de modo que

$$\frac{Y_{1i} - Y_{0i}}{Y_{0i}} = \exp(\beta) - 1.$$

Cuando $\beta$ es menor que, aproximadamente, 0,2, entonces $\exp(\beta) - 1$ se parece lo suficiente a $\beta$ como para que esté justificado referirse a esta última cantidad como un cambio porcentual.[22]

Quizá usted se encuentre con maestros que describen los coeficientes que resultan de una regresión logarítmica-lineal como «puntos logarítmicos». Esta terminología recuerda al auditorio que la interpretación en términos de puntos porcentuales es sólo aproximada. En general, los puntos logarítmicos subestiman el cambio porcentual, es decir,

$$\beta < \exp(\beta) - 1,$$

donde la diferencia entre ambas cantidades aumenta a medida que crece $\beta$. Por ejemplo, cuando $\beta = 0{,}05$, entonces $\exp(\beta) - 1 = 0{,}051$, pero cuando $\beta = 0{,}3$ entonces $\exp(\beta) - 1 = 0{,}35$.

### *Errores típicos de la regresión e intervalos de confianza*

Nuestro tratamiento de la regresión ha ignorado en buena medida el hecho de que nuestros datos provienen de muestras. Como ya indicamos en el apéndice del capítulo primero, las estimaciones resultantes de regresiones muestrales, al igual que las medias muestrales, están sujetas a la varianza muestral. Imaginamos que la relación subyacente cuantificada por la regresión es de carácter fijo y no aleatorio, pero contamos con que se manifiesten diferencias al calcular las estimaciones de esa relación cuando se empleen muestras distintas extraídas de la misma población. Supongamos que queremos definir la relación entre los ingresos de los graduados universitarios y el tipo de universidad al que asistieron. Es poco probable que dispongamos de datos sobre la totalidad de la población de graduados. Por lo tanto, en la práctica, se trabaja con muestras extraídas de la población de

[22] La interpretación en términos de cambio porcentual de modelos de regresión logarítmicos no requiere establecer una comparación de resultados potenciales, pero resulta más sencilla de explicar de este modo.

interés. (Incluso aunque tuviéramos información completa sobre la población estudiantil de un año, estudiantes diferentes conformarán esa población en otros años.) El conjunto de datos analizado para alcanzar las estimaciones de las tablas (2.2)-(2.5) se basa en una de tales muestras. Nos gustaría cuantificar la varianza muestral correspondiente a estas estimaciones.

Como sucedía con la media muestral, la varianza muestral de un coeficiente de regresión se mide por medio de su error típico. El apéndice del capítulo 1 explica que el error típico de una media muestral es:

$$SE(\overline{Y}_n) = \frac{\sigma\gamma}{\sqrt{n}}.$$

El error típico de una pendiente estimada mediante regresión simple $\hat{\beta}$ circunflejo tiene un aspecto similar y se puede escribir como

$$SE(\hat{\beta}_n) = \frac{\sigma_c}{\sqrt{n}} \times \frac{1}{\sigma_c},$$

donde $\sigma_e$ es la desviación típica de los residuos de la regresión, y $\sigma_X$ la desviación típica del regresor $X_i$.

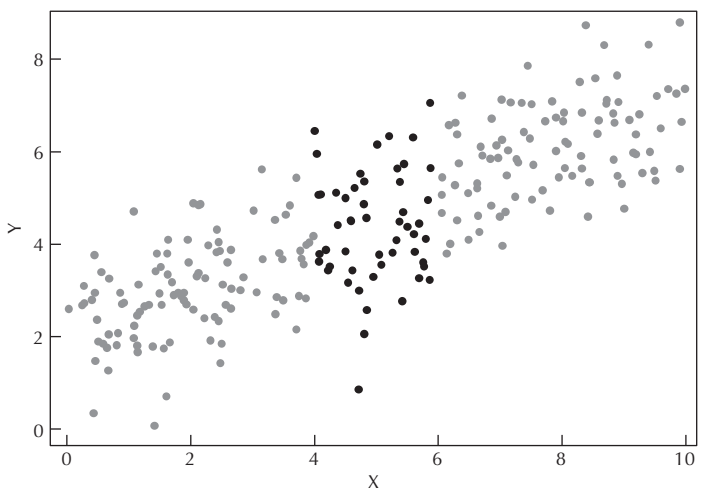



**Figura 2.2.** La varianza en *X* es buena.

Igual que el error típico de una media muestral, los errores típicos de una regresión decrecen cuando aumenta el tamaño de la muestra. Los errores típicos crecen (es decir, las estimaciones de la regresión resultan menos precisas) cuando los residuos presentan una varianza grande. Esto no debe sorprender, porque una gran varianza en los residuos significa que la regresión no da un buen ajuste. Por otra parte, la variabilidad de los regresores es beneficiosa: cuando se incrementa $\sigma_X$ se hace más precisa la pendiente estimada. Esto se aprecia en la figura 2.2, la cual revela que al añadir variabilidad a $X_i$ (en concreto, al añadir las observaciones representadas en gris) se contribuye a realzar la pendiente que vincula $Y_i$ con $X_i$.

La fórmula de la anatomía de la regresión para las regresiones múltiples se puede trasladar a los errores típicos. En un modelo multivariado como este:

$$Y_i = \alpha + \sum_{k=1}^{k} \beta_k X_{ki} + e_i,$$

el error típico para la pendiente $k$-ésima, $\hat{\beta}_k$, es

$$SE(\hat{\beta}_k) = \frac{\sigma_e}{\sqrt{n}} \times \frac{1}{\sigma_{\tilde{X}_k}}, \tag{2.15}$$

donde $\sigma_{\tilde{X}_k}$ es la desviación típica de $\tilde{X}_{ki}$, el residuo de una regresión de $X_{ki}$ sobre el resto de regresores. Añadir controles conlleva dos efectos opuestos sobre $SE(\hat{\beta}_k)$. La varianza de los residuos ($\sigma_e$ en el numerador de la fórmula del error típico) cae cuando se añaden a la regresión variables explicativas para predecir $Y_i$. Por el contrario, la desviación típica de $\tilde{X}_{ki}$ en el denominador de la fórmula del error típico es menor que la desviación típica de $X_{ki}$, lo que incrementa el error. El añadido de variables explicativas explica parte de la variación de otros regresores, y esta variación se elimina en virtud de la anatomía de la regresión. La interrelación entre estos cambios hacia arriba o hacia abajo puede conducir tanto a una mejora como a un empeoramiento de la precisión.

Los errores típicos que se calculan por medio de la ecuación (2.15) se consideran hoy día pasados de moda, y no se suelen mostrar en público. Esa fórmula antigua da por supuesto que la varianza de los residuos no guarda relación con los regresores o, como dicen los maestros, los residuos son *homocedásticos*. Cuando los residuos son

homocedásticos, las estimaciones que resultan de la regresión son estadísticamente eficientes. Sin embargo, esta condición podría no satisfacerse, y por eso los iniciados prefieren hoy día emplear una fórmula más complicada llamada de los *errores típicos robustos.*

La fórmula de los errores típicos robustos se puede escribir como

$$RSE(\hat{\beta}) = \frac{1}{\sqrt{n}} \frac{V(X_{ki} e_i)}{(\sigma^2_{\tilde{X}_k})^2}. \tag{2.16}$$

Los errores típicos robustos tienen en cuenta la posibilidad de que la curva de regresión ajuste mejor o peor para distintos valores de $X_i$, circunstancia en la que se dice que los residuos son *heterocedásticos.* Pero si al final resultara que los residuos fueran homocedásticos, entonces el numerador de la fórmula robusta se simplificaría:

$$V(\tilde{X}_{ki} e_i) = V(\tilde{X}_{ki}) V(e_i) = \sigma^2_{\tilde{X}_k} \sigma^2_e.$$

En este caso, las estimaciones de $RSE(\hat{\beta})$ deberían parecerse a las de $SE(\hat{\beta})$, porque los errores típicos teóricos son entonces idénticos. Pero si los residuos son realmente heterocedásticos, entonces las estimaciones de $RSE(\hat{\beta})$ suelen brindar una medida más acertada (y normalmente algo mayor) de la varianza muestral.[23]

[23] La distinción entre los errores robustos y los errores típicos ya pasados de moda para las estimaciones por regresión, se parece a la distinción (comentada en el apéndice del capítulo 1) entre los estimadores de los errores típicos para la diferencia de dos medias que usan estimaciones comunes o independientes de $\sigma Y_2$ para la varianza de los datos de los grupos de tratamiento y de control.

3

# Variables instrumentales

Kwai Chang Caine: De una sola acción tú extraes todo un universo.

*Kung Fu*, primera temporada, episodio 1

## *Nuestro camino*

El control estadístico por medio de regresiones podría no conducir a estimaciones convincentes de efectos causales. Por fortuna, hay otros caminos que también llevan a *lo demás permanece constante*. Como en los experimentos aleatorios, las fuerzas de la naturaleza, entre las que incluimos la naturaleza humana, a veces manipulan los tratamientos de un modo que hace innecesarios los controles. Es poco frecuente que tales fuerzas sean la única fuente de variación de los tratamientos, pero este obstáculo es fácil de soslayar. El método de las *variables instrumentales* (VI) aprovecha la existencia de procedimientos aleatorios, parciales o incompletos, sean de origen natural o provocados por los investigadores. Ilustramos de tres maneras esta idea crucial. La primera evalúa una innovación educativa de Estados Unidos (escuelas *charter*) con un análisis elemental de VI que aprovecha el carácter aleatorio de los sorteos de admisión. Una segunda aplicación de las VI se centra en cuál es el mejor modo de atajar la violencia de género, y muestra que las VI se pueden utilizar para analizar experimentos de campo que asignan a los sujetos de manera aleatoria unos tratamientos que son libres de rechazar. La tercera aplicación explora

los efectos a largo plazo del hecho de haberse criado en familias grandes o pequeñas. Este caso ilustra el método de mínimos cuadrados en dos etapas (MC2E), un refinamiento del método de VI que constituye una de nuestras herramientas más poderosas.

## 3.1 El dilema *charter*

Entrevistadora: ¿Te han contado papá y mamá lo del sorteo?
Daisy: El sorteo, ¿eso no era que la gente jugaba para ganar dinero?

*Waiting for Superman*, 2010

El documental *Waiting for Superman* [*Esperando a Supermán*] narra la historia de los solicitantes de escolarización en escuelas *charter* de Nueva York y California. Su emisión intensificó un debate ya ferviente acerca de las políticas educativas en Estados Unidos. La película argumenta que las escuelas *charter* brindan las mejores opciones para estudiantes pobres y pertenecientes a minorías que de otro modo habrían permanecido en las escuelas públicas urbanas ordinarias, donde pocos alcanzan la excelencia y muchos terminan abandonando los estudios.

Las escuelas *charter*, también conocidas (sobre todo en California) como escuelas semiautónomas, son centros educativos públicos que funcionan con un grado de autonomía considerablemente mayor que el del resto de colegios públicos estadounidenses. Un concesionario independiente (normalmente de carácter privado, o una organización gestora sin ánimo de lucro) firma un contrato que le reconoce el derecho a gestionar un colegio público durante un periodo limitado, con la renovación del acuerdo condicionada a la obtención de buenos resultados. Estas escuelas tienen libertad para elaborar sus currículos y métodos escolares. Muchas escuelas *charter* amplían el tiempo de enseñanza con jornadas lectivas más largas y prolongando las actividades durante los fines de semana y el verano. La diferencia quizá más importante, y sin duda más controvertida, entre las escuelas *charter* y las públicas tradicionales radica en que la plantilla docente de las primeras no suele estar afiliada a sindicatos. En contraste, la mayoría del personal docente de las escuelas públicas de las gran-

des ciudades tiene contratos supervisados por los sindicatos, en los que se regulan los salarios y las condiciones de trabajo, a veces de un modo muy detallado. Estos contratos pueden mejorar las condiciones laborales del profesorado, pero también dificultan recompensar a los buenos docentes, o despedir a los malos.

Entre las escuelas que aparecen en *Waiting for Superman* se encuentra el centro KIPP LA College Prep,[1] uno de los más de 140 afiliados al Programa Conocimiento Es Poder, conocido como KIPP por sus siglas en inglés (Knowledge Is Power Program). Los centros KIPP son representativos del principio de «sin excusas», un modelo de escuelas *charter* muy imitado y que pone el énfasis en la disciplina y el buen comportamiento, con jornadas lectivas largas, un calendario escolar ampliado, procesos selectivos para contratar profesorado y con gran atención a las habilidades tradicionales de lectura y matemáticas. El programa KIPP lo iniciaron en Houston (Tejas) y en la ciudad de Nueva York en 1995 unos veteranos de Teach for America ['Enseñe por Estados Unidos'], un programa que reclutó a miles de recién graduados de las universidades más selectas de Estados Unidos para que se dedicaran a la enseñanza en distritos escolares con malos resultados académicos. Hoy día, la red KIPP cuenta con un alumnado cuya extracción es en un 95% negra o hispana, y más del 80% de los estudiantes KIPP son lo bastante pobres como para cumplir los requisitos del programa del gobierno federal para becas de comedor escolar.[2]

El debate sobre la reforma de educación en Estados Unidos suele centrarse en la brecha educativa, una manera breve de referirse a una incómoda y profunda diferencia de calificaciones dependiendo de la raza y procedencia étnica. El alumnado negro o hispano suele rendir muy por debajo del blanco (anglosajón) o asiático en los exámenes oficiales. La cuestión de qué políticas habría que aplicar para reducir estas diferencias tan grandes y persistentes entre grupos sociales suele recibir dos respuestas distintas. La primera dirige la mirada hacia

[1] Es decir, Escuela Preparatoria (preuniversitaria) KIPP de Los Ángeles. *(N. de la T.)*

[2] Véanse los detalles del KIPP en el libro de Jay Mathews *Work Hard. Be Nice*, Algonquin Books, 2009. Teach for America fue en 2012 el mayor empleador de graduados en 55 campus universitarios de Estados Unidos, desde la Universidad del Estado de Arizona hasta Yale.

las escuelas para mejorar los resultados; la segunda reclama un cambio social más amplio, con el argumento de que no parece probable que estos bajos rendimientos los corrijan los centros educativos por sí solos. KIPP suele estar en el centro del debate, por su especialización en alumnado procedente de minorías, y sus defensores subrayan que los estudiantes KIPP que no son blancos sacan en los exámenes notas marcadamente superiores a las obtenidas por el alumnado no blanco de las escuelas próximas. Los escépticos responden que el éxito aparente del KIPP refleja que este programa atrae a familias cuyos hijos cuentan ya de por sí con más probabilidades de éxito:

> Los estudiantes KIPP, como colectivo, acceden al programa KIPP con calificaciones superiores a las habituales en las escuelas de las que proceden. [E]l profesorado nos decía que, o bien trataban a estudiantes superiores a sus compañeros, o que las familias más proclives a esta iniciativa y a entrar en KIPP eran las más motivadas y con una formación más completa.[3]

Este comentario suscita la pregunta clave de si *ceteris* es *paribus* cuando se compara a los estudiantes KIPP con el alumnado de otros centros públicos.

### *Una lotería*

La primera escuela KIPP de Nueva Inglaterra fue un centro de secundaria en la ciudad de Lynn, Massachusetts, justo al norte de Boston. Un viejo refrán dice: «Lynn, Lynn, ciudad del pecado de donde saldrás del todo cambiado». Pero no es que haya muchas cosas saliendo de Lynn hoy en día, con o sin pecado. En tiempos fue un gran centro de manufactura de calzado, pero Lynn ha destacado recientemente por sus altos índices de desempleo, criminalidad y pobreza. La mayoría de los estudiantes de las escuelas públicas de Lynn no son blancos, y en 2009 más de las tres cuartas partes de ellos cumplían los requisitos para acceder a becas de comedor escolar. El nivel de pobreza es incluso mayor en la cohorte de estudiantes de Lynn que acceden al

[3] Martin Carnoy, Rebecca Jacobsen, Lawrence Mishel y Richard Rothstein, *The Charter School Dust-Up: Examining Evidence on Student Achievement*, Economic Policy Institute Press, 2005, p. 58.

quinto curso en KIPP. Los centros *charter* urbanos suelen matricular a muchos estudiantes pobres y negros, pero la escuela KIPP de Lynn es excepcional en cuanto a que acoge una proporción elevada de muchachos hispanos con un conocimiento limitado de la lengua inglesa.

La escuela KIPP de Lynn arrancó a medio gas en 2004, con menos solicitudes que plazas disponibles. Un año después había más peticiones que plazas, aunque por poco. Sin embargo, a partir de 2005 la demanda se aceleró, con más de 200 solicitudes para unas 90 plazas en quinto curso cada año. La ley de Massachusetts exige que en estos casos las plazas se asignen por sorteo. Estos sorteos son mucho más que un curioso adorno institucional, porque nos permiten descifrar el enigma causal de las escuelas *charter*. Nuestra herramienta VI recurre a los sorteos de admisión para aprovechar un experimento aleatorio que sucede de manera natural.

En realidad, la decisión de estudiar en un centro *charter* nunca es totalmente aleatoria: incluso entre los solicitantes ocurre a veces que tras conseguir plaza deciden acudir a otra escuela, mientras que algunos no agraciados en el sorteo acaban encontrando alguna vía de acceso distinta. Sin embargo, las comparaciones entre los solicitantes que *consiguen* una plaza y los que no, como resultado de los sorteos de admisión aleatorios, deberían presentar naturalezas lo bastante semejantes como para considerar que juntamos manzanas con manzanas. Si se supone que la única diferencia como resultado de ganar el sorteo estriba en que se altera la probabilidad de matricularse en el centro *charter* (supuesto denominado *restricción de exclusión*), entonces el método VI convierte los efectos de la asignación aleatoria de plazas en estimaciones causales del efecto de estudiar en una escuela *charter*. En concreto, las estimaciones por medio de VI captan los efectos causales sobre aquellos sujetos que se matriculan en el centro KIPP tras ganar el sorteo, pero que no habrían accedido al mismo de otro modo. Como se explica a continuación, este grupo se conoce como el conjunto que *accede* por el sorteo KIPP.

El maestro Joshway y sus colaboradores recopilaron datos de solicitantes KIPP en Lynn desde el otoño de 2005 hasta el otoño de 2008.[4]

[4] Joshua D. Angrist *et al.*, «Inputs and Impacts in Charter Schools: KIPP Lynn», *American Economic Review Papers and Proceedings*, vol. 100, número 2, mayo de 2010, páginas 239-243, y Joshua D. Angrist *et al.*, «Who Benefits from KIPP?», *Journal of Policy Analysis and Management*, vol. 31, número 4, otoño de 2012, páginas 837-860.

Algunos solicitantes tenían derecho a ahorrarse el sorteo: quienes tuvieran ya hermanos o hermanas en el centro tenían la admisión garantizada (en buena medida). Algunas solicitudes quedaban descartadas desde el principio (las de personas demasiado mayores para cursar estudios de secundaria, por ejemplo). De las 446 solicitudes de entrada en quinto curso que se sometieron a asignación aleatoria en los cuatro sorteos KIPP de 2005 a 2008, 303 (68%) fueron agraciadas. Pero puede sorprender que, sin embargo, un número considerable de estas personas al final no se matriculó en septiembre. Algunas se habían mudado, mientras que otras prefirieron al final una escuela próxima en su casa. De entre quienes consiguieron una plaza, 221 (73%) cursaron estudios en KIPP durante el año académico siguiente. A la vez, un puñado de los solicitantes no agraciados en el sorteo (en torno a un 3,5%) encontraron algún otro modo de acceder a KIPP (hubo algunos casos de solicitudes no agraciadas que consiguieron plaza en fechas posteriores, o en los siguientes sorteos). La figura 3.1 resume esta información relevante.

Los sorteos KIPP asignan las plazas al azar. Esta asignación aleatoria debería equilibrar las características demográficas de los solicitantes agraciados y no agraciados. El equilibro de condiciones entre agraciados parece realmente bueno, como se puede apreciar en el apartado A de la tabla 3.1. La primera columna reproduce, como base de referencia, las características demográficas y las calificaciones en educación primaria de todos los estudiantes de quinto grado de las escuelas públicas de Lynn. Las columnas segunda y tercera contienen los promedios de los agraciados en los sorteos KIPP y la diferencia de las medias entre agraciados y perdedores, que muestra que ambos grupos contienen proporciones iguales de negros, de hispanos o de gente con recursos lo bastante limitados para tener derecho a beca de comedor escolar.

Un rasgo especialmente importante de la tabla 3.1 es que permite comprobar que este equilibrio existe antes de aplicarse el tratamiento, en concreto las calificaciones que tuvieron en cuarto curso (antes del acceso a KIPP) los solicitantes de acceso al quinto curso, lo que se denomina «notas de partida» en la tabla. Como suele suceder en los estudios sobre resultados académicos, estas puntuaciones se han *estandarizado* mediante la sustracción de la media y dividiendo por la desviación típica de las notas en una población de referencia, en este

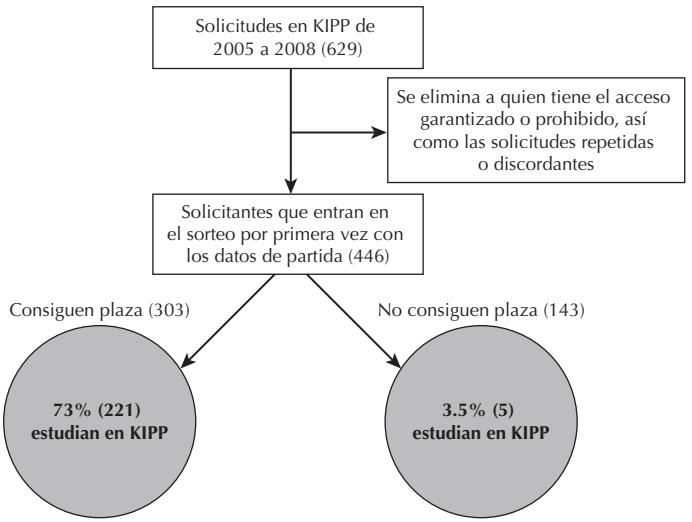



**Figura 3.1.** Datos de solicitudes y matrículas de los sorteos KIPP en Lynn
*Nota:* Se muestra entre paréntesis el número de solicitantes del programa Knowledge is Power Program (KIPP).

caso la población de estudiantes de cuarto curso de Massachusetts. Al estandarizar se obtienen calificaciones que se miden en unidades definidas por la desviación típica de la población de referencia. Como en muchas de las localidades más pobres de Massachusetts, las notas medias en matemáticas en Lynn se sitúan unas tres décimas de desviación típica por debajo de la media del estado. Este nivel de puntuación se representa como $-0{,}3\sigma$ (como en el apéndice de los capítulos 1 y 2, la desviación típica se designa mediante la letra griega $\sigma$, «sigma»). Las diferencias en datos de partida entre agraciados y no agraciados en los sorteos KIPP son pequeñas y estadísticamente no significativas, como se ve en la columna (3) de la tabla 3.1, y lo más probable es que se deban al puro azar.

Las dos últimas columnas de la tabla 3.1 muestran los promedios para los estudiantes de quinto curso que se matricularon en la escuela KIPP de Lynn, junto a las diferencias entre los solicitantes KIPP que se matricularon y los que no lo hicieron. Como la matriculación no está asignada al azar, las diferencias entre matriculados y no matriculados podrían presentar cierto sesgo de selección: los ganadores del sorteo que eligieron acudir a otros centros podrían tener menos

## Tabla 3.1. Análisis de los sorteos KIPP

| | Estudiantes de quinto curso en escuelas públicas de Lynn | Solicitantes KIPP | | | |
|---|---|---|---|---|---|
| | | Agraciados en el sorteo KIPP de Lynn | Agraciados frente a no agraciados | Asistieron a KIPP | Estudiaron en KIPP frente a otros |
| | (1) | (2) | (3) | (4) | (5) |
| **Apartado A. Datos de partida** | | | | | |
| Hispanos | .418 | .510 | –.058 | .539 | .012 |
| | | | (.058) | | (.054) |
| Negros | .173 | .257 | .026 | .240 | –.001 |
| | | | (.047) | | (.043) |
| Mujeres | .480 | .494 | –.008 | .495 | –.009 |
| | | | (.059) | | (.055) |
| Con beca de comedor | .770 | .814 | –.032 | .828 | .011 |
| | | | (.046) | | (.042) |
| Nota de partida en matemáticas | –.307 | –.290 | .102 | –.289 | .069 |
| | | | (.120) | | (.109) |
| Nota de partida en lengua | –.356 | –.386 | .063 | –.368 | .088 |
| | | | (.125) | | (.114) |
| **Apartado B. Resultados** | | | | | |
| Asistieron a KIPP | .000 | .787 | .741 | 1.000 | 1.000 |
| | | | (.037) | | — |
| Nota en matemáticas | –.363 | –.003 | .355 | .095 | .467 |
| | | | (.115) | | (.103) |
| Nota en lengua | –.417 | –.262 | .113 | –.211 | .211 |
| | | | (.122) | | (.109) |
| Tamaño de la muestra | 3.964 | 253 | 371 | 204 | 371 |

*Notas:* Esta tabla describe los datos de partida de los estudiantes de quinto curso en Lynn, y refleja los efectos estimados de la oferta de plazas del Knowledge Is Power Program (KIPP) sobre los solicitantes de Lynn. Las medias aparecen en las columnas (1), (2) y (4). La columna (3) presenta las diferencias entre las personas agraciadas y no agraciadas en el sorteo. Se trata de coeficientes procedentes de regresiones que incluyen controles para grupos de riesgo en la forma de variables binarias que marcan el año y el curso de la solicitud, así como la existencia de hermanos solicitantes. La columna (5) muestra las diferencias entre los estudiantes de KIPP y los solicitantes que no asistieron a KIPP. Los errores típicos constan entre paréntesis.

interés por la enseñanza que quienes aceptaron la oportunidad de acceder a KIPP. Este es el tipo de sesgo de selección que suelen aducir los críticos del sistema KIPP. Sin embargo, resulta que las diferencias que se aprecian en la columna (5) son pequeñas, y ninguna de ellas se acerca a la significancia estadística, lo que indica que el sesgo de selección tal vez no sea importante en este contexto.

La mayoría de las solicitudes piden el acceso en quinto curso, un año antes de que den comienzo los estudios secundarios reglados, pero hay quien solicita entrar en sexto. Aquí consideramos los efectos de estudiar en KIPP sobre los resultados académicos en los exámenes que se hacen al final del curso siguiente al de la solicitud. Estas notas son las del final del quinto curso para quienes pidieron el ingreso cuando estaban en cuarto, y las notas finales de sexto para quienes estaban en quinto cuando cursaron la solicitud. La muestra resultante, que incluye 371 solicitudes, omite a personas jóvenes que pidieron el ingreso antes de terminar el tercer curso, así como a otros solicitantes que no aportaron notas y datos de partida.[5]

El apartado B de la tabla 3.1 ilustra que los solicitantes KIPP que consiguieron plaza obtuvieron calificaciones estandarizadas cercanas a cero, es decir, en torno a la media del estado. Como los solicitantes KIPP parten de notas de cuarto curso que en promedio caen $0{,}3\sigma$ por debajo de la media del estado, este logro a nivel de media estatal podría parecer impresionante. En contraste, la nota media de aquellos que no consiguieron una plaza ronda $-0{,}36\sigma$, un poco por debajo del punto de partida en cuarto curso.

Como el sorteo es aleatorio, la diferencia entre 0 y –0,36 que consta en la columna (3) corresponde a un efecto causal medio: la oferta de una plaza en el colegio KIPP de Lynn mejora la nota en matemáticas en $0{,}36\sigma$, una ganancia grande (el efecto de obtener un puesto en KIPP sobre las notas en lectura resulta menor, aunque positivo, y no es estadísticamente significativo). Mencionemos como nota técnica que el análisis que se presenta aquí es algo más complicado que una

[5] Como se indicó en el capítulo 1, la atrición (falta de datos) es motivo de preocupación incluso en los experimentos aleatorios. La clave para garantizar la integridad de un experimento aleatorio en el que faltan datos reside en que se dé la misma probabilidad de que falten datos en los grupos de tratamiento y de control. En la muestra KIPP usada para elaborar la tabla 3.1, agraciados y no agraciados tienen, de hecho, la misma probabilidad de contar con todos los datos requeridos.

simple comparación de las medias, aunque la idea subyacente sea la misma. Los resultados de la columna (3) proceden de efectuar regresiones de las notas sobre una variable binaria que marca la oferta de plaza en KIPP, junto con otras variables binarias que indican el año y el curso (quinto o sexto), y la presencia o no de hermanos o hermanas entre los solicitantes. Estas variables de control son necesarias porque la probabilidad de ganar el sorteo varía de un año a otro y de un curso a otro, y es mucho mayor para hermanos. Las variables de control que hemos empleado describen grupos de estudiantes (a veces denominados *conjuntos de riesgo*) que tienen una probabilidad constante de ganar el sorteo.[6]

¿Qué nos dice un efecto de 0,36$\sigma$ sobre las consecuencias de cursar estudios en la escuela KIPP de Lynn? El estimador VI convierte los efectos de conseguir plaza en KIPP en efectos de cursar estudios en KIPP. En este caso, la *variable instrumental* (o, para abreviar, el «instrumento») es una variable binaria que marca a los solicitantes que consiguieron una plaza. En general, un instrumento cumple los requisitos siguientes:

(i) El instrumento ejerce un efecto causal sobre la variable cuya incidencia tratamos de definir, en este caso la matriculación en KIPP. Por motivos que pronto quedarán claros, este efecto causal se denomina *la primera etapa*.

(ii) El instrumento se asigna al azar, o de un modo «tan bueno como al azar», en el sentido de que se hace de una manera que no guarda relación con las variables omitidas que nos gustaría controlar (en este caso, variables como el contexto familiar o las motivaciones). Esto se conoce como *supuesto de independencia*.

(iii) Para terminar, la lógica VI requiere una *restricción de exclusión*. La restricción de exclusión describe un canal único a través del cual el instrumento ejerce sus efectos sobre los resultados. Aquí la restricción de exclusión equivale a afirmar que el diferencial de notas de 0.36$\sigma$ entre ganadores y perdedores del sorteo sólo se puede atribuir a la diferencia de 0,74 en asistencia a KIPP entre ganadores y perdedores que consta en la columna (3) de la tabla 3.1 (en la parte superior del apartado B).

[6] El apartado 3.3 detalla el papel de las covariables en la estimación VI.

El método VI recurre a tres supuestos para caracterizar una reacción en cadena que conduce desde el instrumento hasta los resultados académicos. El primer eslabón de esta cadena causal (la primera etapa) conecta la asignación aleatoria de oferta de plazas con la asistencia efectiva a la escuela KIPP, mientras que el segundo eslabón (que es el que nos interesa) vincula asistencia a KIPP con resultados. En virtud del supuesto de independencia y de la restricción de exclusión, el producto de estos dos eslabones genera el efecto de la oferta de plaza sobre las notas en los exámenes:

$$\begin{aligned}&\textit{Efecto de la oferta de plaza sobre la puntuación} = \\ &\quad = ([\textit{Efecto de la oferta de plaza sobre la asistencia}] \times \\ &\qquad \times [\textit{Efecto de la asistencia sobre la puntuación}]).\end{aligned}$$

Si se reordena la ecuación, el efecto causal de cursar estudios en KIPP sería:

$$\textit{Efecto de la asistencia sobre la puntuación} = \frac{[\textit{Efecto de la oferta de plaza sobre la puntuación}]}{[\textit{Efecto de la oferta de plaza sobre la asistencia}]} \qquad (3.1)$$

El resultado es $0{,}48\sigma$, tal como consta a la izquierda de la figura 3.2.

Es fácil resumir la lógica que subyace a la ecuación (3.1): se admite que la oferta de plaza en KIPP afecta a los resultados académicos tan sólo a través de la asistencia al centro KIPP. El ofrecimiento de plaza incrementa la tasa de asistencia a un centro KIPP en 75 puntos porcentuales (un factor 0,74, para ser precisos), así que si se multiplican los efectos de la oferta de plaza en los resultados por un factor 4/3 (que es aproximadamente igual a 1/0,74) se genera el efecto de la asistencia. Este ajuste corrige el hecho de que aproximadamente una cuarta parte de las personas que consiguen plaza en KIPP elige acudir a otros centros de enseñanza, mientras que hay algunas que se matriculan en KIPP sin haber ganado el sorteo de plazas.

Las columnas (4) y (5) de la tabla 3.1 presentan una estimación alternativa del efecto de acudir a un centro KIPP. La columna (4) refleja las notas medias de los estudiantes KIPP, mientras que la (5)

muestra el contraste entre los estudiantes KIPP y cualquier otra persona de todo el conjunto de solicitantes. Las diferencias de la columna (5) no tienen en cuenta el sorteo aleatorio para conseguir plazas, y proceden de una regresión de las notas en matemáticas después de la matriculación, sobre una variable binaria que marca la asistencia a KIPP, junto a los mismos controles que se usaron para construir las diferencias entre agraciados y no agraciados de la columna (3). La variación en cuanto a asistencia a KIPP en esta regresión procede en su mayor parte, aunque no del todo, del sorteo. Como la matriculación en KIPP implica tanto una asignación aleatoria como una opción personal (efectuada, por ejemplo, cuando los agraciados eligen no ir a KIPP), las comparaciones entre quienes se matricularon y quienes no lo hicieron podrían estar afectadas de sesgo de selección. Sin embargo, la estimación de la nota en matemáticas de la columna (5) (en torno a 0,47$\sigma$) se acerca mucho a la estimación VI de la figura 3.2, lo que confirma nuestra conjetura anterior, en el sentido de que el sesgo de selección carece de importancia en este caso.

Una mejora de media desviación típica en las notas de matemáticas al cabo de un curso escolar constituye un efecto considerable. Los residentes en Lynn que han tenido la suerte de asistir al centro KIPP, realmente salen cambiados.

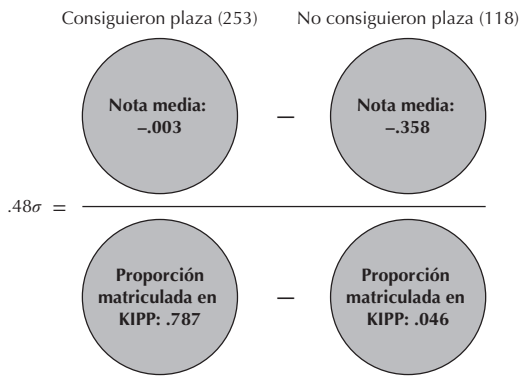



**Figura 3.2.** VI en la escuela: el efecto de estudiar en KIPP sobre las notas en matemáticas.
*Nota:* el efecto de la matriculación en KIPP (Knowledge Is Power Program) descrito en esta figura asciende a 0,48$\sigma$ = 0,355$\sigma$/0,741.

### *ELMT en las escuelas* charter

El sorteo del KIPP brinda un ejemplo de reacción en cadena VI. A los componentes de tales reacciones se les han asignado nombres para que los maestros puedan debatir con eficacia acerca de ellos. Ya hemos indicado que el elemento aleatorio original (en este caso, la oferta de plazas en KIPP) se denomina *variable instrumental*, o instrumento para abreviar. Como hemos visto, el enlace desde el instrumento hasta la variable causal de interés (en este caso, el efecto del sorteo para ofertar plazas sobre la asistencia a escuelas KIPP) se denomina primera etapa, porque es el primer eslabón de la cadena. El efecto directo del instrumento sobre los resultados, que recorre todo el largo de la cadena (en este caso, el efecto del ofrecimiento de plaza sobre las notas), se llama *forma reducida*. Por último, el efecto causal de interés (el segundo eslabón de la cadena) se conoce como el cociente entre la forma reducida y las estimaciones de la primera etapa. Este efecto causal se llama *efecto local medio del tratamiento* (ELMT para abreviar).

Los eslabones de la cadena VI están hechos de diferencias entre valores esperados condicionados, es decir, comparaciones de promedios poblacionales de distintos grupos. En la práctica, los promedios poblacionales se estiman como medias muestrales, normalmente usando datos procedentes de muestras aleatorias. Los datos necesarios son:

- el instrumento, $Z_i$: en este caso, una variable binaria igual a 1 para solicitantes que consiguieron una plaza KIPP al azar (definida sólo para los participantes en el sorteo);
- la variable de tratamiento, $D_i$: en este caso, una variable binaria igual a 1 para quienes asistieron al centro KIPP (por motivos históricos, esto se denomina a veces variable endógena), y
- la variable de resultado, $Y_i$: en este caso, las notas de matemáticas en quinto curso.

Las relaciones clave entre estas variables, es decir, los eslabones de la cadena VI, son parámetros. Por eso los designamos (tal como usted habrá adivinado ya) con letras griegas.

La primera etapa $E[D_i|Z_i = 1] - E[D_i|Z_i = 0]$; llamada $\phi$.

*En el estudio KIPP, $\phi$ («fi») es la diferencia en cuanto a índices de asistencia al centro KIPP entre quienes ganaron y quienes no ganaron el sorteo (igual a 0,74 en la figura 3.2).*

La forma reducida $E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]$; llamada $\phi$.

*En el estudio KIPP, $\phi$ («ro») es la diferencia en cuanto a puntuaciones medias entre solicitantes agraciados y no agraciados en el sorteo (igual a 0,36 en la figura 3.2).*

El efecto local medio del tratamiento (ELMT)

$$\lambda = \frac{\rho}{\phi} = \frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]}; \tag{3.2}$$

ELMT, aquí llamado $\lambda$ («lambda»), es la razón entre la forma reducida y la primera etapa.

*En el estudio KIPP, ELMT es la diferencia de notas entre agraciados y no agraciados dividida entre la diferencia en cuanto a proporción de asistencia al centro KIPP entre agraciados y no agraciados (igual a 0,48 en la figura 3.2).*

Podemos estimar $\lambda$ si sustituimos los cuatro valores esperados poblacionales del segundo miembro de la ecuación (3.2) por sus correspondientes medias muestrales, un estimador que los maestros llaman VI. Sin embargo, en la práctica se suele optar por un método conocido como mínimos cuadrados en dos etapas (MC2E), que se describe en detalle más adelante, en el apartado 3.3. MC2E implementa la misma idea, pero le añade flexibilidad. De un modo u otro, el hecho de que los parámetros se estimen a través de muestras requiere cuantificar su varianza muestral por medio de los errores típicos correspondientes. No le resultará sorprendente saber que hay una fórmula VI para los errores típicos, y que los programas de cálculo econométrico la llevan incorporada. ¡Problema resuelto!

Una cuestión más interesante es la que se refiere al significado de $\lambda$: justo el valor del ELMT de las escuelas *charter*. Puede que el alumnado se beneficíe de manera distinta por acudir a centros KIPP. Para algunas personas, quizá las que cuenten con el apoyo de su en-

torno familiar, la elección entre la escuela KIPP de Lynn o algún otro centro público de la localidad no tenga gran relevancia; el efecto causal de la asistencia a KIPP sobre tales solicitantes sería 0. Pero para otros alumnos la asistencia a KIPP podría tener mucha trascendencia. El ELMT consiste en un promedio de estos efectos causales individuales. En concreto, el ELMT es el efecto causal promedio para el alumnado cuya matriculación en KIPP se debe exclusivamente al hecho de haber ganado el sorteo.

El relato bíblico de la Pascua explica que hay cuatro tipos de muchachos, y lo mismo ocurre con la juventud de hoy. Empecemos por los primeros tres tipos. Los solicitantes como Silvio desean ansiosamente ingresar en KIPP; si no ganan el sorteo, sus madres conseguirán que entren como sea. Los solicitantes como Camila se alegrarán de ir a KIPP si ganan una plaza, pero aceptarán estoicamente el veredicto si no resultan agraciados en el sorteo. Finalmente, tenemos a solicitantes como Normando, abrumados por la perspectiva de unas jornadas escolares largas y una gran cantidad de deberes. En realidad, Normando no quiere ingresar en KIPP y se niega a hacerlo cuando se entera de que ha conseguido una plaza por sorteo. A Normando se le podría denominar *nunca tomador*, porque su elección de escuela no se ve afectada por el sorteo (su papeleta entró en él porque algún trabajador social la introdujo en la urna). En el otro extremo del compromiso KIPP se encuentra Silvio, un *siempre tomador* que siempre ocupará plaza si le dan una, y cuya madre siempre halla el modo de que esto ocurra aunque pierda en el sorteo, quizá colando una solicitud falsificada simulando que cuenta con un hermano agraciado en el sorteo. En el caso de Silvio, la elección de escuela tampoco se ve afectada por el resultado del sorteo.

Camila se matricula en KIPP sólo si consigue plaza en el sorteo, pero acudirá, aunque sea a regañadientes, a otra escuela del vecindario en caso contrario (la madrina de Camila no da abasto y, aunque desea lo mejor para su ahijada, sólo puede jugar las cartas de la baza que le ha tocado en el reparto). Camila pertenece a la categoría de solicitantes que dan al método VI todo su poder, porque el instrumento altera su estado de tratamiento. Cuando su $Z_i = 0$, resulta que $D_i = 0$, y si su $Z_i = 1$ entonces $D_i = 1$. Las estrategias VI dependen de sujetos como Camila, que llamaremos *cumplidores*, un grupo que marcaremos con la variable binaria $C_i$. El término cumplidores (*com-*

*pliers* en inglés) procede del mundo de los experimentos aleatorios. En muchos ensayos aleatorios, como los que se efectúan para valorar medicamentos nuevos, la decisión de cumplir o no con un tratamiento asignado al azar sigue siendo voluntaria y no aleatoria (los sujetos experimentales a los que por azar se ofrece el tratamiento podrían negarse a tomarlo, por ejemplo). Los cumplidores en tales experimentos son quienes toman el tratamiento si se les asigna, pero no en otro caso. En el sorteo de instrumentos es el ELMT el que representa el efecto causal promedio de asistir a un centro KIPP sobre Camila o sobre otros cumplidores que se matriculan en KIPP si, y sólo si, ganan el sorteo. Los métodos VI no aportan información acerca de los siempre tomadores como Silvio, ni sobre los nunca tomadores como Normando, porque el instrumento no guarda relación con su estado de tratamiento.

**Tabla 3.2. Los cuatro tipos de muchachos**

| | | **No agraciados en el sorteo $Z_i = 0$** | |
|---|---|---|---|
| | | No asisten a KIPP $D_i = 0$ | Asisten a KIPP $D_i = 1$ |
| Agraciados en el sorteo $Z_i = 1$ | No asisten a KIPP $D_i = 0$ | Nunca tomadores *(Normando)* | Retadores |
| | Asisten a KIPP $D_i = 1$ | Cumplidores *(Camila)* | Siempre tomadores *(Silvio)*} |

*Nota:* KIPP = Knowledge Is Power Program.

La tabla 3.2 clasifica a sujetos como Silvio, Normando o Camila, pero incluye también un cuarto grupo, el de los *retadores*. Las columnas indican la elección de matriculación que se realiza cuando $Z_i = 0$, mientras que las filas reflejan las elecciones que se efectúan cuando $Z_i = 1$. La tabla cubre todas las posibilidades para cualquier solicitante, y no sólo para aquellos que podemos observar (por ejemplo, para los solicitantes agraciados en el sorteo la tabla incluye también qué opciones habrían elegido en caso de perder). Los nunca tomadores, como Normando, y los siempre tomadores como Silvio, constan en la diagonal principal. Ganen o pierdan, su elección de centro escolar no se ve alterada. Abajo a la izquierda está Camila, que cumple con lo que le ofrece el sorteo y se matricula en KIPP si, y sólo si, resulta agraciada. La primera etapa, $E[D_i|Z_i = 1] - E[D_i|Z_i = 0]$, se basa en estos

solicitantes, y el ELMT refleja los efectos promedio del tratamiento sobre este grupo.

Los retadores de la tabla 3.2 serían aquellos que se matricularían en KIPP si, y sólo si, no se les ofreciera una plaza a través del sorteo. La Biblia se refiere a estos rebeldes como «malvados», pero no hagamos juicios morales. Sí tenemos que darnos cuenta, sin embargo, de que este comportamiento perverso dificulta la interpretación de las estimaciones por medio de VI. Si entre los datos hay tanto retadores como cumplidores, entonces el efecto medio de conseguir una plaza en KIPP podría ser 0 incluso si todo el mundo se beneficiara de la asistencia a tales escuelas. Por suerte es poco probable que se den actitudes retadoras en los sorteos de escuelas *charter*, así como en muchos otros estudios con VI. Por lo tanto, aceptamos que las actitudes retadoras son entre poco frecuentes e inexistentes. Este supuesto de ausencia de retadores se denomina *monotonía*, palabra que significa que el instrumento actúa sobre los solicitantes tan sólo en un sentido.

Hemos explicado que las variables instrumentales se pueden interpretar como los agentes iniciadores de una cadena causal en las que un instrumento, $Z_i$, altera la variable de interés, $D_i$, que a su vez repercute en los resultados, $Y_i$. El concepto de población cumplidora ligada a cada instrumento desempeña una función central en la interpretación de esta reacción en cadena. El teorema del ELMT afirma que para cualquier instrumento asignado de manera aleatoria con una primera etapa no nula, y que satisfaga tanto la monotonía como la restricción de exclusión, el cociente entre la forma reducida y la primera etapa es igual al ELMT, el efecto causal promedio del tratamiento sobre los cumplidores.[7] Recordemos (apartado 1.1) que $Y_{1i}$ denota el resultado para el sujeto $i$ cuando se activa el tratamiento, mientras que $Y_{0i}$ es el resultado para la misma persona cuando el tratamiento no se aplica. Si se usa esta notación, junto con los parámetros definidos más arriba, el ELMT se puede escribir como:

[7] Este teorema se debe a Guido W. Imbens y Joshua D. Angrist, «Identification and Estimation of Local Average Treatment Effects», *Econometrica*, vol. 62, número 2, marzo de 1994, páginas 467-475. La distinción entre cumplidores, siempre tomadores y nunca tomadores se detalla en Joshua D. Angrist, Guido W. Imbens y Donald B. Rubin, «Identification of Causal Effects Using Instrumental Variables», *Journal of the American Statistical Association*, vol. 91, número 434, junio de 1996, páginas 444-455.

$$\lambda = \frac{\rho}{\phi} = E[Y_{1i} - Y_{0i}|C_i = 1].$$

Sin supuestos más estrictos, como que el efecto causal es constante para todos los sujetos (este sería el modelo descrito por la ecuación [1.3] del capítulo 1), el ELMT no tiene por qué revelar efectos causales sobre nunca tomadores ni sobre siempre tomadores.

No debería sorprendernos que una variable instrumental no siempre sea útil para conocer los efectos sobre personas cuyo estado de tratamiento no se puede alterar mediante la manipulación del instrumento. La buena noticia es que la población de cumplidores constituye el grupo acerca del cual queremos saber algo. En el ejemplo KIPP los cumplidores son los estudiantes que posiblemente asistieran a centros KIPP si se ampliara la red de esas escuelas y se ofrecieran así más plazas en el sorteo, quizá como consecuencia de la apertura de centros nuevos en la misma zona. En Massachusetts, donde el número de plazas en centros *charter* está limitado por ley, las consecuencias de tal ampliación constituyen un tema de política educativa que está a la orden del día.

Los investigadores y los políticos se interesan a veces por los efectos causales promedio para toda la población tratada, y no sólo en el ELMT. Este efecto causal promedio se denomina *efecto del tratamiento sobre los tratados* (TST). El TST se escribe como $E[Y_{1K} - Y_{0K}|D_i = = 1]$. Por norma hay dos maneras de recibir tratamiento, es decir, de activar la variable $D_i$. Una es recibir tratamiento con independencia de si el instrumento se activa o no. Como hemos explicado, esta es la historia de Silvio, siempre tomador. El resto de la población tratada consta de cumplidores a los que de manera aleatoria se asigna $Z_i = 1$. En el estudio KIPP, la muestra tratada incluye cumplidores que consiguieron una plaza (como Camila), y siempre tomadores (como Silvio) que asistieron a centros KIPP con independencia de todo lo demás. La población de cumplidores a los que se ofreció plaza al azar es representativa de la población de todos los cumplidores (incluidos aquellos cumplidores que perdieron en el sorteo y acudieron a escuelas públicas), pero los efectos sobre los siempre tomadores no tienen por qué ser los mismos que sobre los cumplidores. Cabría imaginar, por ejemplo, que Silvio fuera siempre tomador porque su madre tuviera la convicción de que KIPP le cambiará la vida. El efecto causal que experimente será por tanto mayor que

el de solicitantes sujetos al tratamiento, pero menos motivados, es decir, los cumplidores tratados.

Como la población sujeta a tratamiento incluye a siempre tomadores, el ELMT y el TST no suelen coincidir. Además, ninguno de estos dos efectos causales promedio tiene por qué mantenerse a lo largo del tiempo, o en contextos diferentes (por ejemplo, en centros *charter* con menos solicitantes pertenecientes a minorías étnicas). La cuestión de si una estimación causal concreta posee valor predictivo para otros tiempos, lugares y personas, aparte de las representadas en el estudio que la produjo, se denomina *validez externa*. Para estimar la validez externa de un resultado los maestros tienen que plantearse por qué una estimación particular de ELMT es grande o pequeña. Parece creíble, por ejemplo, que KIPP mejore los resultados académicos porque la fórmula KIPP proporcione un ambiente educativo en el que muchos estudiantes (pero quizá no todos) aprendan con más facilidad. Un alumnado especialmente brillante e independiente podría no mejorar en KIPP. Para explorar la validez externa de un ELMT concreto se puede usar un solo instrumento que genere estimaciones sobre estudiantes de diversos tipos como, por ejemplo, los que tienen notas de partida mayores o menores. También pueden considerarse otros instrumentos que afecten a categorías distintas de cumplidores, un tema al que volveremos en el apartado 3.3. Como con las estimaciones resultantes de los experimentos aleatorios, el mejor indicio a favor de la validez externa de las estimaciones VI procede de la comparación entre los ELMT que resultan de aplicar el mismo tratamiento, o tratamientos similares, a poblaciones distintas.

## 3.2 Contra el abuso

La policía recibió llamadas desde la mansión de O. J. Simpson en Los Ángeles al menos nueve veces en el curso de su matrimonio con Nicole Brown Simpson. Pero quien fuera estrella de la liga estadounidense de fútbol americano, apodado *The Juice* (*El Zumo*), sólo fue arrestado una vez, en 1989, cuando no formuló alegaciones frente a una acusación de violencia de género por un episodio que llevó a Nicole al hospital. Simpson pagó una multa pequeña, prestó servicios comuni-

tarios y tuvo que acudir a la consulta de un psiquiatra de su elección. El fiscal del caso en 1989, Robert Pingle, constató que Nicole no había sido muy proclive a colaborar con las autoridades en fechas posteriores a la considerable paliza recibida. Cinco años después Nicole Brown Simpson y su compañero Ronald Goldman fueron asesinados por un intruso desconocido que muchas personas creen que era el ex esposo de Nicole, O. J.[8]

¿Cómo tendría que responder la policía ante la violencia de género? Como en el caso de Nicole Brown Simpson, las víctimas de estos abusos suelen resistirse a presentar cargos. Arrestar a los maltratadores sin la cooperación de las víctimas podría no llevar a nada, y contribuir a agravar una situación ya de por sí mala. Para muchos observadores y para no pocos responsables policiales, las agencias de servicios sociales parecen mejor preparadas para atajar la violencia de género. A la vez, los abogados de las víctimas están preocupados porque el hecho de que no se arreste a los maltratadores revela una tolerancia social hacia actos violentos que, si sucedieran entre extraños, despertarían una respuesta legal contundente.

En el tumulto de un debate político acalorado, el alcalde y jefe de la policía de Minneapolis emprendió un experimento pionero a comienzos de la década de 1980. El Experimento de Violencia de Género de Minneapolis (conocido como MDVE, de *Minneapolis Domestic Violence Experiment*) se diseñó para diagnosticar la eficacia de detener a los maltratadores.[9] El plan de investigación del MDVE incluía tres tratamientos: arresto, orden de abandonar el inmueble durante ocho horas (separación), e intervención de profesionales que podría incluir mediación por parte de los policías presentes en la escena (asesoramiento). El proyecto establecía la aplicación de uno de estos tres tratamientos de manera aleatoria cada vez que los funcionarios de la

[8] Simpson fue absuelto de asesinato en un juicio penal, pero se lo consideró responsable de las muertes en un juicio civil. Luego publicó un libro titulado *If I Did It: Confessions of the Killer* [*Si lo hubiera hecho: confesiones del asesino*], Beaufort Books, 2007. Nuestro relato de las visitas repetidas de la policía al hogar de los Simpson se basa en Sara Rimer, «The Simpson Case: The Marriage; Handling of 1989 Wife-Beating Case Was a "Terrible Joke", Prosecutor Says» [«El caso Simpson: el matrimonio; el fiscal afirma que el desarrollo del caso de la agresión a la esposa fue una "broma atroz"»], *The New York Times*, 18 de junio de 1994.

[9] El análisis original del MDVE aparece en Lawrence W. Sherman y Richard A. Berk, «The Specific Deterrent Effects of Arrest for Domestic Assault», *American Sociological Review*, vol. 49, número 2, abril de 1994, páginas 261-272.

policía de Minneapolis se encontraran ante una situación que satisficiera los criterios experimentales (en concreto, que hubiera motivos suficientes para creer que un compañero o esposo hubiera atacado con violencia a su compañera durante las últimas cuatro horas). Se excluían los casos en que había riesgo de muerte o de heridas graves (delito de lesiones). Tanto el sospechoso como su víctima tenían que estar presentes a la llegada de los policías. El resultado primario que quería observarse en MDVE era la reincidencia dentro de los seis meses posteriores al primer tratamiento aleatorio.

El dispositivo para la asignación aleatoria de tratamiento en MDVE era un mazo de formularios con códigos de colores asigandos al azar, para cada una de los tres posibles tratamientos: arresto, separación o asesoramiento. Los policías que se encontraban una situación que satisficiera los criterios experimentales tenían que actuar según el color del primer formulario del mazo (el situado más arriba). Los policías que participaron en el experimento se ofrecieron voluntarios para ello y, por tanto, se esperaba que aplicaran el diseño experimental. A la vez, todo el personal implicado en el estudio sabía que era poco realista, e incluso inapropiado, seguir estrictamente el protocolo de aleatorización.

**Tabla 3.3. Tratamientos asignados y aplicados en MDVE**

| | **Tratamiento aplicado** | | | |
|---|---|---|---|---|
| Tratamiento asignado | | Indulgencia | | |
| | Arresto | Consejo | Separación | Total |
| Arresto | 98.9 (91) | 0.0 (0) | 1.1 (1) | 29.3 (92) |
| Consejo | 17.6 (19) | 77.8 (84) | 4.6 (5) | 34.4 (108) |
| Separación | 22,84 (26) | 4.4 (5) | 72.8 (83) | 36.3 (114) |
| Total | 43.4 (136) | 28.3 (89) | 28.3 (89) | 100.0 (314) |

*Notas:* Esta tabla presenta los porcentajes y números absolutos de la distribución de tratamientos asignados y aplicados en el Experimento de Minneapolis sobre Violencia de Género (MDVE, *Minneapolis Domestic Violence Experiemnt*). Las tres primeras columnas presentan los porcentajes dentro de cada fila. La última columna contiene los porcentajes dentro de esa columna. Los números absolutos de casos constan entre paréntesis.

En la práctica, los policías solían desviarse de aplicar el tratamiento que señalaba el color del formulario en el momento del incidente.

Había casos en que se arrestaba al sospechoso aunque la asignación aleatoria indicara separación o consejo. La mayoría de los arrestos en estas condiciones se producían cuando un sospechoso intentaba agredir a un policía, cuando una víctima pedía el arresto de manera insistente, o cuando ambas partes habían resultado heridas. Algunas desviaciones surgieron porque los policías se olvidaron de llevar los formularios. Como resultado de estas desviaciones del protocolo experimental, el *tratamiento aplicado* no fue estrictamente aleatorio. Esto puede verse en la tabla 3.3, que muestra los tratamientos asignados y los aplicados. En casi todos los casos en que correspondía arrestar al sospechoso, el arresto se produjo (91 de 92 asignaciones), pero muchos casos en los que tocaba separación o asesoramiento terminaron también en arresto.

El contraste entre el arresto, que solía implicar una noche en el calabozo, y otras alternativas más indulgentes genera los resultados más interesantes y controvertidos de MDVE. La tabla 3.3, por tanto, combina los dos tratamientos que no implicaban arresto bajo el encabezamiento «indulgencia». La asignación aleatoria ejercía un efecto grande, pero no determinante, sobre la probabilidad de que el sospechoso de agresión fuera tratado con indulgencia: la probabilidad de que se aplicara un trato indulgente a un caso al que se hubiera asignado tal tratamiento ascendía a 0,797 ({insertar aquí la fórmula con los cocientes de sumas que aparece en la línea 5 de la página 118}); mientras que los casos a los que no tocaba indulgencia (o sea, salió arresto) recibían ese trato suave con una probabilidad de 0,011 (1/92). Como la indulgencia no se aplicó estrictamente al azar, MDVE podría parecer un experimento fallido. Pero los métodos VI, sin embargo, permiten rectificar con facilidad este fallo.

### *Cuando ELMT es el efecto sobre quien recibe tratamiento*

El esquema ELMT tiene su motivación en una analogía entre el método VI y los experimentos aleatorios. El método VI nos permite captar el efecto causal del tratamiento sobre los sujetos tratados, a pesar de que los participantes en los experimentos hayan tomado decisiones que vulneran la aleatoriedad, como en el caso MDVE. De hecho, el uso de VI suele ser necesario en tales experimentos.

Un análisis ingenuo de los datos MDVE basado en los tratamientos aplicados resultaría engañoso.

Un análisis de MDVE basado en el tratamiento administrado sería engañoso porque los casos en los que los funcionarios de policía deberían haber aplicado indulgencia a los sospechosos de maltrato y, en efecto así lo hicieron, constituyen un subconjunto no aleatorio entre todos los casos a los que tocó indulgencia. La comparación entre quienes recibieron un trato indulgente y los que no está, por tanto, contaminada por un sesgo de selección. Los maltratadores que fueron arrestados, aunque aleatoriamente les hubiera tocado indulgencia, solían ser especialmente agresivos o estar muy nerviosos. Esta fuente de sesgo de selección se elimina si se usa el tratamiento asignado de manera aleatoria como variable instrumental para el tratamiento efectivamente aplicado.

Como siempre, la reacción en cadena VI empieza con la primera etapa.[10] La primera etapa en MDVE es la diferencia entre la probabilidad de ser tratado con indulgencia cuando el tratamiento era deindulgencia, y la probabilidad de ser tratado con indulgencia cuando el tratamiento era de arresto. Indiquemos mediante $Z_i$ si indulgencia fue el tratamiento aleatorio, y denotemos como $D_i$ si efectivamente se aplicó indulgencia. La primera etapa de este procedimiento es:

$$E[D_i|Z_i = 1] - E[D_i|Z_i = 0] = 0{,}797 - 0{,}011 = 0{,}786,$$

una diferencia grande, pero aun así lejos de la diferencia igual a la unidad que obtendríamos si el protocolo se hubiera seguido con rigor.

Por desgracia, la violencia de género suele ser reincidente, como constata el hecho de que la policía recibiera una segunda llamada para intervenir por el mismo motivo en el 18% de las direcciones de la muestra MDVE. Lo más importante desde el punto de vista de la investigación MDVE es que la reincidencia fue mayor entre los sospechosos a los que se asignó indulgencia que entre los sospechosos a los que se asignó arresto. Esto se deduce si se calcula el efecto de la asignación aleatoria de indulgencia sobre una variable de resultado,

[10] Nuestro análisis VI de los datos MDVE se basa en Joshua D. Angrist, «Instrumental Variables Methods in Experimental Criminological Research: What, Why and How», *Journal of Experimental Criminology*, vol. 2, número 1, abril de 2006, páginas 23-44.

$Y_i$, que indica la existencia de al menos un episodio de sospecha de abuso tras el tratamiento:

$$E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0] = 0{,}211 - 0{,}097 = 0{,}114. \qquad (3.3)$$

Dado que el índice global de reincidencia asciende al 18%, esta diferencia estimada de 11 puntos porcentuales es sustancial.

En los ensayos aleatorios con cumplimiento imperfecto del protocolo, donde el tratamiento asignado difiere del tratamiento aplicado, los efectos de la asignación aleatoria como el calculado en la ecuación (3.3) se denominan *efectos de intención de tratamiento* (IT). El análisis de IT capta el efecto causal de asignar un tratamiento, pero ignora el hecho de que algunos de los sujetos a los que se asignó un tratamiento indulgente terminaron siendo arrestados. Como el efecto IT no tiene en cuenta este incumplimiento del protocolo, su valor resulta demasiado pequeño comparado con el efecto causal promedio de la indulgencia sobre todos los sujetos que efectivamente recibieron ese trato. Sin embargo, es fácil resolver este problema: si se dividen los efectos IT entre la diferencia de índices de cumplimiento del protocolo entre los grupos de tratamiento y de control, se captura el efecto causal de la indulgencia sobre los cumplidores tratados de ese modo como consecuencia del experimento.

Dividir las estimaciones IT de un experimento aleatorio entre la diferencia correspondiente en cuanto a cumplimiento del protocolo constituye otro caso de VI en acción: reconocemos IT como la forma reducida de un instrumento asignado de manera aleatoria, en concreto la asignación aleatoria de indulgencia. Como hemos visto, muchos sospechosos de maltrato a los que se asignó indulgencia acabaron arrestados de todos modos. La regresión de una variable binaria que marque si se aplicó indulgencia sobre otra binaria que indique la asignación aleatoria de indulgencia constituye la primera etapa correspondiente a esta forma reducida. La cadena causal VI empieza con la asignación aleatoria de tratamiento, continúa con el tratamiento aplicado y finalmente afecta a los resultados.

La estimación ELMT que resulta de los datos MDVE es impresionante: 0,114/0,786 = 0,145, un efecto grande de la indulgencia, incluso cuando se compara con las estimaciones IT. Hay que remarcar que, a pesar de que los policías presentes en la escena fueran muy

selectivos a la hora de decidir si seguían o no el protocolo experimental, esta estimación de ELMT probablemente represente una buena medida del efecto causal del tratamiento aplicado.

Como siempre, la interpretación causal del ELMT depende en parte de la relevante restricción de exclusión, que requiere que la variable de tratamiento de interés constituya la única vía a través de la cual el tratamiento puede afectar a los resultados. En el caso MDVE, la reacción en cadena VI comienza con el color del formulario que usa el funcionario de policía. La restricción de exclusión requiere aquí que el color del formulario asignado al azar afecte a la reincidencia tan sólo a través de la decisión de arrestar o de tratar con indulgencia a los sospechosos de maltrato. Este supuesto parece razonable, aún más cuando tanto los maltratadores como sus víctimas ignoran que están participando en un estudio experimental.

¿Son realmente necesarias las modestas complicaciones que implica el análisis VI? Supongamos que analizamos los datos MDVE utilizando la información de los tratamientos aplicados, ignorando la naturaleza no aleatoria de unas decisiones que tendrían que haber seguido un protocolo de asignación al azar. El análisis resultante compararía la reincidencia entre quienes fueron y no fueron tratados con indulgencia, sin más complejidades ni ajustes:

$$E[Y_i|D_i = 1] - E[Y_i|D_i = 0] = 0{,}216 - 0{,}129 = 0{,}087.$$

El efecto estimado ahora resulta inferior al resultante del análisis VI en casi 15 puntos porcentuales.

El capítulo 1 muestra que, sin asignación aleatoria, las comparaciones entre sujetos tratados y no tratados es igual al efecto causal de interés más el sesgo de selección. El sesgo de selección que contaminaría un análisis ingenuo de los datos MDVE es la diferencia en cuanto a reincidencia potencial (es decir, en $Y_{0i}$) entre maltratadores que recibieron y no recibieron indulgencia. Aunque buena parte de la variación en la indulgencia aplicada se produjo por asignación aleatoria, los funcionarios en la escena de los hechos también tomaron decisiones discrecionales. Los maltratadores que fueron arrestados a pesar de que el azar les había asignado indulgencia solían ser especialmente violentos o estar particularmente nerviosos, mientras que los sospechosos de los casos en los que la policía cumplió la asig-

nación de indulgencia solían ser más sumisos. En otras palabras, los agresores a los que se aplicó indulgencia tenían, ya de por sí, menos probabilidades de reincidir en cualquier caso. El sesgo de selección resultante conduce a que el cálculo basado en el tratamiento aplicado infravalore el efecto de la indulgencia. En contraste con el estudio KIPP (tratado en el apartado 3.1), en este caso el sesgo de selección es relevante.

El análisis VI de los datos MDVE elimina el sesgo de selección y refleja los efectos causales promedio sobre los cumplidores (en este caso, el efecto de tratar con indulgencia a los maltratadores en incidentes en los que la policía procedió a aplicar una asignación aleatoria de ese tratamiento). Un rasgo interesante e importante del estudio MDVE es que los incumplimientos del protocolo se produjeron casi sólo en un sentido. Cuando la asignación aleatoria indicaba arresto, la policía arrestaba sin pensarlo dos veces (con una única excepción entre 92 casos). En contraste, en más del 20% de los casos en que tocaba indulgencia se aplicó arresto.

Esta asimetría en el cumplimiento de la aleatoriedad en los casos de indulgencia significa la ausencia casi total de siempre tomadores en el estudio MDVE. En nuestro análisis VI de los datos MDVE, los siempre tomadores serían sospechosos de maltrato destinados a recibir indulgencia con independencia del tratamiento asignado. El tamaño de este grupo viene dado por la probabilidad de recibir indulgencia cuando toca arresto, en este caso tan sólo 1/92. Como indicamos en el apartado 3.1, cualquier población de tratamiento es la unión de dos grupos, el conjunto de cumplidores a los que se asigna tratamiento de manera aleatoria y el conjunto de siempre tomadores. Si no hay siempre tomadores, todos los tratados son cumplidores, en cuyo caso ELMT coincide con TST:

$$\lambda = E[Y_{1i} - Y_{0i}|C_i = 1] = E[Y_{1i} - Y_{0i}|D_i = 1].$$

Si se aplica la propiedad de inexistencia de siempre tomadores en los datos MDVE, se ve que ELMT es igual al efecto causal promedio de la indulgencia sobre los tratados con indulgencia. En concreto, la estimación TST que surge de MDVE contrasta la reincidencia de los tratados con indulgencia ($E[Y_{1i}|D_i = 1]$) con los índices que observaríamos en un mundo contrafactual en el que se hubiera arrestado

a los maltratadores que realmente fueron tratados con indulgencia ($E[Y_{0i}|D_i = 1]$). Esta importante simplificación del contexto habitual del ELMT surge en cualquier análisis VI carente de siempre tomadores, incluyendo muchos otros experimentos aleatorios en los que el incumplimiento del protocolo se produce siempre en un sentido determinado. Cuando algunos de los sujetos a los que el azar asignó tratamiento resultan no tratados, pero no hay nadie que haya sido asignado aleatoriamente al grupo de control que reciba tratamiento, entonces los métodos VI que usan intenciones de tratamiento aleatorias como un instrumento para el tratamiento aplicado, dan como resultado el TST.[11]

Una nota final sobre la importancia de la buena econometría: es difícil sobreestimar el impacto del estudio MDVE sobre la aplicación de la ley en Estados Unidos. Ahora se procede de manera rutinaria a arrestar a los agresores incluso en casos leves de violencia de género. En muchos estados se ha instaurado la obligatoriedad del arresto en casos sospechosos de violencia de género.

Pequeño Saltamontes: Maestro, el caso O.J. sucedió una década después del MDVE. El diseño experimental pionero MDVE no salvó la vida de Nicole Brown y Ron Goldman.

Maestro Joshway: El cambio social transcurre pausado, Pequeño Saltamontes. Y aquellos analistas originales de MDVE comunica-

[11] Este resultado teórico procede de Howard S. Bloom, «Accounting for No-Shows in Experimental Evaluation Designs», *Evaluation Review*, vol. 8, número 2, abril de 1984, páginas 225-246. La interpretación del ELMT a la luz de los resultados de Bloom aparece en Imberns y Angrist, «Identification and Estimation», *Econometrica*, 1994. Véase también el apartado 4.4.3 de Joshua D. Angrist y Jörn-Steffen Pischke, *Mostly Harmless Econometrics: An Empiricists Companion*, Princeton University Press, 2009. Un ejemplo procedente de nuestro campo de la econometría del trabajo es Job Training Partnership Act (JTPA). El experimento JTPA asignó al azar la oportunidad de participar en un programa financiado de formación laboral a escala federal. Alrededor del 60% de las personas a las que se ofreció formación recibieron los servicios de JTPA, pero no hubo miembros del grupo de control que lo hicieran. Un análisis VI de JTPA usando el tratamiento asignado como instrumento para el tratamiento aplicado, capta el efecto de la formación sobre las personas que la cursaron. Para más detalles véase Larry L. Orr *et al.*, *Does Training for the Disadvantaged Work? Evidence from the National JTPA Study*, Urban Institute Press, 1996.

ron las estimaciones ingenuas basadas en los tratamientos aplicados, junto con los efectos de intención de tratamiento. Las estimaciones de mi estudio VI de 2006 son mucho mayores.

Pequeño Saltamontes: ¿Habrían salvado la vida Nicole y Ron si los analistas anteriores hubieran usado variables instrumentales?

Maestro Joshway: Hay cosas que nunca podremos saber.

### 3.3 La bomba poblacional

¿Control de la población, o abocados a la extinción?

Paul Ehrlich, 1968

La población mundial creció desde 3.000 millones hasta 6.000 millones entre 1960 y 1999; así pues, se duplicó en 39 años, y esto sucedió en la mitad del tiempo que se necesitó para pasar de 1.500 millones a 3.000 millones. Y bastaron tan sólo una docena de años para llegar a los siguientes 1.000 millones. Pero en la demografía contemporánea existe consenso entorno a que el crecimiento de la población se ha frenado de manera espectacular. Las proyecciones basadas en la tasa de fertilidad actual apuntan a que para duplicarse de nuevo se necesitarán al menos cien años, si es que jamás se duplica. Una estimación muy citada prevé un máximo de población de 9.000 millones en 2070.[12] A pesar de todo lo que se escribe actualmente sobre desarrollo sostenible, la bomba poblacional ha quedado desactivada. ¡Qué alivio!

El estudio del modo en que el crecimiento de la población afecta a los niveles de vida tiene tanto una vertiente «macro» como una «micro». Las raíces de la macrodemografía se alargan hasta el estudioso inglés del siglo xviii Thomas Malthus, que defendía que la población crece cuando aumenta la producción de alimentos, de manera que los incrementos de productividad no se reflejan en una mejora

[12] Véase David Lam, «How the World Survived the Population Bomb: Lessons from 50 Years of Extraordinary Demographic History», *Demography*, vol. 48, número 4, noviembre de 2011, páginas 1231-1262, y Wolfgang Lutz, Warren Sanderson y Sergei Scherbov, «The End of World Population Growth», *Nature*, vol. 412, número 6846, 2 de agosto de 2001, páginas 543-545.

del nivel de vida. La triste conclusión malthusiana predice una existencia anclada permanentemente en el nivel de subsistencia para la mayoría de la población. La historia ha contradicho en repetidas ocasiones esta visión pesimista del crecimiento económico, pero eso no ha impedido que ganara predicamento entre los más recientes agoreros del Apocalipsis. El libro superventas de Paul Ehrlich del año 1968 titulado *The Population Bomb* lanzó a la fama un guión malthusiano que preveía una hambruna masiva e inminente en India. La población de India se ha triplicado desde entonces, a la vez que el nivel de vida ha experimentado un crecimiento notable.[13]

La economía ha colocado una microlente sobre la relación entre el tamaño de las familias y el nivel de vida. Aquí la atención se centra en la capacidad de unidades familiares de distintos tamaños para mantener un nivel de vida holgado. De hecho, cabría esperar que el aumento de la familia incrementara la pobreza y redujera el grado de formación (más bocas que alimentar implica menos recursos para cada una) y eso es lo que muestran las correlaciones simples. Una racionalización teórica más elaborada de esta poderosa relación procede del trabajo del difunto Gary Becker y sus colaboradores. Estos estudios introdujeron el concepto de «equilibrio entre cantidad y calidad», la idea de que la reducción del tamaño de las familias aumenta la inversión parental en los hijos. Por ejemplo, los padres con menos hijos pueden vigilar más de cerca la salud de su descendencia, e invertir más en su educación.[14]

Desde un punto de vista político, la idea de que es fundamental reducir el tamaño de las familias para incrementar el nivel de vida ha animado a las agencias internacionales y muchos gobiernos a fomentar, y a veces incluso imponer, unidades familiares reducidas. China abrió el camino con la controvertida política del hijo único, introducida en 1979. Otras iniciativas agresivas promovidas por los gobiernos

[13] Se debate hasta qué punto ha subido el nivel de vida en India. Aun así, los estudios suelen coincidir en que las condiciones han mejorado de manera espectacular desde 1970 (véase, por ejemplo, Angus Deaton, *The Great Escape: Health, Wealth, and the Origins of Inequality*, Princeton University Press, 2013).

[14] Gary S. Becker y H. Gregg Lewis, «On the Interaction between the Quantity and Quality of Children», *Journal of Political Economy*, vol. 81, número 2, parte 2, marzo-abril de 1973, páginas S279-288, y Gary S. Becker y Nigel Tomes, «Child Endowements and the Quantity and Quality of Children», *Journal of Political Economy*, vol. 84, número 4, parte 2, agosto de 1976, páginas S143-S162.

con el objeto de una planificación familiar incluyen un programa de esterilización forzosa en India, y la promoción pública de la planificación familiar en México y en Indonesia. Hacia 1990 el 85% de la población del mundo en vías de desarrollo vivía en países donde los gobiernos consideraban la fertilidad elevada como una de las causas principales que perpetúan la pobreza.[15]

Es difícil encontrar argumentos en contra de la correlación negativa entre el tamaño promedio de las familias y los indicadores de desarrollo, como la escolaridad. Pero ¿hay una relación causal entre el tamaño familiar y el nivel de formación de los hijos? El desafío a la hora de abordar esta cuestión reside, como siempre, en la *paribus*icidad del *ceteris*. La fertilidad depende en su mayor parte de las decisiones que toman los progenitores.[16] No debe sorprender, por tanto, que las mujeres con familias grandes difieran en muchos aspectos de las que tienen familias pequeñas: suelen tener menos formación, por ejemplo. Y la descendencia de madres con un nivel cultural más bajo suele estar, a su vez, peor formada. Las diferencias tan marcadas en cuanto a características observables entre familias de distintos tamaños disparan las señales de alarma sobre posibles sesgos de selección. Como las mujeres con distinto número de hijos son tan diferentes desde el punto de vista observacional, debemos reconocer la posibilidad de que haya otras diferencias no observadas pero importantes relacionadas con el tamaño familiar.

Como siempre, la solución ideal para un problema de variables omitidas sería la asignación aleatoria. En este caso el experimento podría organizarse del modo siguiente. (i) Tome una muestra de familias con un solo hijo. (ii) Distribuya un hijo adicional, al azar, entre estas familias. (iii) Espere 20 años y recoja datos sobre los logros académicos de los primogénitos que recibieron un hermano adicional,

---

[15] John Bongaarts, «The Impact of Population Policies: Comment», *Population and Developement Review*, vol. 20, número 3, septiembre de 1994, páginas 616-620.

[16] Cabría pensar que esto sólo es cierto en sociedades con acceso a métodos anticonceptivos modernos, como la píldora o la moneda de céntimo (que se sostiene entre las rodillas en caso de necesidad). Pero la demografía revela que, incluso sin acceso a anticonceptivos modernos, los progenitores ejercen un grado notable de control sobre la fertilidad. Por ejemplo, en un extenso cuerpo de trabajo Ansley Coale documentó el descenso espectacular de la fertilidad marital en la Europa de los siglos XIX y XX (véase http://opr.princeton.edu/archive/pefp/). Esta tendencia, continuada desde entonces en la mayor parte del mundo, se denomina la *transición demográfica*.

y de los que no. Por supuesto, no es de esperar que presenciemos la ejecución de un experimento así en un futuro próximo. No obstante, los maestros avezados pueden encontrar fuentes de variación que revelen si hay conexión causal entre el tamaño familiar y la escolaridad sin tener que recurrir a un experimento real.

Esto nos conduce de nuevo a preguntarnos de dónde vienen los niños. Como la mayoría de nuestros lectores ya sabe, los bebés humanos son distribuidos entre los hogares por unas aves de largos picos y patas llamadas cigüeñas (aunque es un mito que dejen caer los niños por las chimeneas, dado que las chimeneas tienen compuertas internas que impiden la entrega de un niño vivo por este sistema). La entrega se materializa nueve meses después de que una mujer, a la que llamaremos «madre», haya declarado su intención de tener un hijo. Las cigüeñas son insensibles a los deseos de los varones (a no ser que tales deseos los canalice una mujer intermediaria), así que nos concentraremos en el experimento desde el punto de vista de las madres y del mayor de sus hijos o hijas.

El experimento que tenemos en mente consiste en añadir hijos a hogares que ya tienen uno. El primogénito será nuestro sujeto experimental. El desafío econométrico consiste en generar una variación en el tamaño familiar de estos sujetos de un modo «tan bueno como si fuera al azar». Por desgracia, la Asociación de Cigüeñas Matronas rechaza toda asignación aleatoria, porque la considera contra natura. Aun así, las cigüeñas a veces generan cierta variación aleatoria en los tamaños de las familias al repartir más de un bebé de una sola vez, en forma de mellizos o gemelos (a consecuencia del hecho de que las cigüeñas son muy grandes y los bebés muy pequeños, así que a veces estas aves se llevan más de uno, sin darse cuenta, al sacar los bebés del almacén). El hecho de que la llegada de mellizos o gemelos induzca un experimento sobre el tamaño familiar se reconoció por primera vez en un estudio pionero de Mark Rosenzweig y Kenneth Wolpin, quienes recurrieron a una pequeña muestra de casos para investigar la disyuntiva cantidad y calidad en India.[17]

Explotaremos el experimento de los mellizos utilizando una muestra grande de Israel, analizada en un estudio sobre la disyuntiva

[17] Mark R. Rosenzweig y Kenneth I. Wolpin, «Testing the Quantity-Quality Fertility Model: The Use of Twins as a Natural Experiment», *Econometrica*, vol. 48, número 1, enero de 1980, páginas 227-240.

cantidad y calidad por el maestro Joshway, con sus colegas Victor Lavy y Analia Schlosser (abreviado el «estudio ALS»).[18] Israel constituye un caso interesante porque su población es muy diversa, incluyendo muchas personas nacidas en países en vías de desarrollo y en familias grandes. Casi la mitad de la población judía israelí es de origen europeo, mientras que la otra mitad es oriunda de Asia o África. En Israel vive también una cantidad notable de árabes, aunque los datos sobre israelíes no judíos son menos completos que los disponibles sobre los judíos. Un rasgo atractivo de la muestra judía israelí consiste en que, aparte de tener más diversidad étnica y familias mayores que las que existen en la mayoría de los países desarrollados, se dispone de información sobre las correspondientes familias de origen, incluyendo las edades y sexos de los hermanos. Esta combinación de datos tan poco frecuente constituye las bases de la estrategia empírica ALS.

Nos centramos aquí en un grupo de primogénitos adultos que conforman una muestra aleatoria de varones y mujeres nacidos de madres con al menos dos vástagos. Estos primogénitos tienen al menos un hermano o hermana más joven, pero muchos tienen dos, o más. Consideremos una familia en la que el segundo nacimiento traiga un solo bebé. En promedio, esas familias tienen 3,6 hijos. Si el segundo parto es doble, entonces el tamaño medio de la familia aumenta en 0,32, es decir, en casi un tercio de hijo. ¿Por qué el nacimiento de mellizos o gemelos incrementa el tamaño de la familia en un salomónico hijo fraccional? Muchas parejas israelíes querrían tener tres o cuatro hijos; el tamaño de sus familias no se ve seriamente afectado por un parto múltiple, porque pretenden tener más de dos hijos, de todos modos. Pero, por otra parte, algunas familias están contentas con sólo dos hijos. Este último grupo se ve forzado a aumentar el tamaño familiar de dos a tres cuando la cigüeña les trae mellizos. Ese diferencial de un tercio de hijo causado por la llegada de los mellizos refleja una diferencia en cuanto a probabilidades: la probabilidad de tener un tercer hijo se incrementa desde aproximadamente 0,7 cuando el segundo parto es simple, hasta la certeza si ese parto es múltiple. La cifra 0,3 se debe a que la diferencia entre la probabilidad 1 y la probabilidad 0,7 es 0,3.

[18] Joshua D. Angrist, Victor Lavy y Analia Schlosser, «Multiple Experiments for the Casual Link between the Quantity and Quality of Children», *Journal of Labor Economics*, vol. 28, número 4, octubre de 2010, páginas 773-824.

Una simple regresión del curso académico más alto completado por los primogénitos adultos sobre el tamaño de la familia muestra que cada hermano o hermana extra implica una reducción de alrededor de un cuarto de año de escolaridad (estos resultados proceden de un modelo que incorpora controles de edad y sexo). Por otra parte, como muestra el estudio ALS, aunque los primogénitos adultos cuyos siguientes hermanos son mellizos hayan crecido en familias más grandes, no tienen menos estudios que los primogénitos adultos procedentes de familias en las que el siguiente hermano nació solo. La comparación de escolaridad entre primogénitos cuyos hermanos siguientes son mellizos o no lo son constituye la forma reducida para una estimación VI que usa los partos múltiples como un instrumento para el tamaño familiar.

Las estimaciones VI se construyen a partir del cociente entre la forma reducida y las estimaciones de la primera etapa, de manera que una forma reducida nula indica de manera inmediata que el efecto causal del número de hermanos que llegan en el parto siguiente es cero. El hecho de que la forma reducida correspondiente al instrumento parto múltiple y las estimaciones VI asociadas estén cerca de cero va en contra de la idea de que proceder de una familia más grande reduce el nivel de formación de los hijos. En otras palabras, el experimento de los mellizos no genera indicios a favor de la hipótesis de que exista una disyuntiva entre cantidad-calidad.

Los partos múltiples ejercen un efecto marcado en el tamaño de las familias, pero los experimentos con mellizos o gemelos no son perfectos. Como la Asociación de Cigüeñas Matronas rechaza toda asignación aleatoria, hay un cierto desequilibrio en la incidencia de los mellizos. Los partos múltiples son más frecuentes entre madres de más edad, o entre mujeres de ciertos grupos étnicos o raciales. Esto puede dar pie a un sesgo de variables omitidas en nuestro análisis del experimento de los mellizos, sobre todo si algunas de las características que fomentan el parto múltiple fueran difíciles de observar y controlar.[19] Por suerte, hay un segundo experimento de fertilidad que aporta datos acerca de la disyuntiva cantidad-calidad.

[19] En muestras más recientes, el uso del parto múltiple como instrumento se ve comprometido también por la proliferación de la fertilización *in vitro*, un tratamiento contra la infertilidad. Las madres que recurren a la fertilización *in vitro*, que incrementa mucho la frecuencia de partos múltiples, tienden a tener más edad y más formación que otras madres.

En muchos países la fertilidad se ve afectada por la composición sexual de hermanos y hermanas. En primer lugar, es frecuente que padres y madres deseen hijos varones, una preferencia especialmente fuerte en algunos lugares de Asia. A los progenitores de Europa, América e Israel parece importarles poco si sus hijos son varones o mujeres. Al contrario, muchas parejas desearían tener un surtido de descendientes de los dos sexos: las familias cuyos primeros dos hijos son ambos varones o ambos mujeres son más propensos a tener un tercero. Como el sexo de los recién nacidos se asigna de manera esencialmente aleatoria (la mitad de los nacimientos son de niños y la otra mitad de niñas y, en ausencia de aborto selectivo, poco puede hacerse para cambiar esto), las preferencias de los padres a favor de una descendencia variada en cuanto a sexos genera instrumentos basados en esta circunstancia.

Los primogénitos adultos israelíes cuyo hermano o hermana siguiente es del sexo opuesto crecieron en hogares que en promedio tenían 3,60 hijos. Pero los primogénitos cuyo siguiente hermano o hermana era de su mismo sexo, tienen familias con 3,68 hijos. Dicho de otro modo, la primera etapa para primogénitos israelíes con siguiente hermano del mismo sexo asciende a 0,08. Como en el caso de la primera etapa cuando los hermanos siguientes son mellizos, esta diferencia refleja cambios en la probabilidad de tener hijos causada por un instrumento. En este caso la variable instrumental es una variable binaria que vale 1 para las familias cuyos dos primeros hijos son del mismo sexo, y 0 para las familias formadas por un niño y una niña. La primera etapa para el instrumento «igualdad de sexos» es más pequeña que la que aparece en el instrumento «parto múltiple», pero el número de familias en las que los dos primeros hijos son del mismo sexo es mucho mayor que el de familias con mellizos o gemelos. Casi la mitad de las familias con al menos dos hijos tienen, o bien dos varones, o bien dos mujeres en los dos primeros nacimientos. En contraste, sólo un 1% de las madres tiene mellizos o gemelos. El instrumento igualdad de sexos entre hermanos tiene otra ventaja sobre el instrumento parto múltiple, en el sentido de que no guarda relación con características de la madre como la edad a la que da a luz o la raza (como muestran tanto el estudio ALS como otro anterior del maestro Joshway y William Evans).[20]

[20] Joshua D. Angrist y William Evans, «Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size», *American Economic Review*, vol. 88, número 3, junio de 1988, páginas 450-477.

Resulta que el nivel educativo de los primogénitos adultos israelíes no está afectado por la distribución de sexos de sus hermanos. Por ejemplo, el curso más alto completado en promedio por los primogénitos es igual a 12,6 tanto cuando los hermanos son del mismo sexo como cuando no.[21] En consecuencia, la forma reducida para el instrumento igualdad de sexos resulta igual a cero y, por tanto, también lo son las estimaciones VI correspondientes. Como en el experimento de los partos múltiples, los cambios de fertilidad causados por diferencias en la diversidad de sexo de los hijos no aportan indicios a favor de la hipótesis de una disyuntiva entre cantidad-calidad.

La restricción de exclusión requerida para una interpretación causal de las estimaciones VI cuando el instrumento es la igualdad de sexos de los hijos afirma que la distribución de sexos de los hermanos influye en los logros de los adultos sólo en la medida en que altere el tamaño de la familia. ¿Es posible que la diversidad sexual entre los dos primeros hermanos repercuta en los resultados escolares por otras vías? Parece probable que dos niños o dos niñas compartan dormitorio durante más tiempo que si son de distinto sexo, por ejemplo, y hermanos o hermanas del mismo sexo se benefician de más posibilidades de intercambiar ropa. Estas ventajas familiares podrían hacer que los hogares con dos descendientes del mismo sexo se sientieran un poco más ricos, un sentimiento que podría acabar incrementando la inversión parental en la escolaridad de los hijos.

¿Hay algún modo de poner a prueba la restricción de exclusión? No de manera directa pero, como suele suceder, se pueden encontrar ciertos indicios al respecto. Hay algunas madres con poca probabilidad de que la diversidad sexual de los hijos incida en su fertilidad. Por ejemplo, en una muestra israelí, las mujeres religiosas que planean tener tres o más hijos ejercen de siempre tomadoras para el instrumento de igualdad de sexos. Por otro lado, las mujeres con mayor formación, la mayoría de las cuales planea tener familias más pequeñas, son nunca tomadoras si su comportamiento en cuanto a fertilidad no se ve alterado por la composición sexual de los hijos. Como la fertilidad de las siempre tomadoras y de las nunca tomadoras no cambia con la distribución de sexos entre los hijos, cualquier relación entre el instrumento igualdad de sexos y los resultados pro-

[21] El número 12 corresponde al último año de la educación secundaria. *(N. del E.)*

cedentes de muestras con pocas mujeres cumplidoras podría indicar violaciones de la restricción de exclusión subyacente.

Podemos expresar esta idea de manera más formal si usamos la notación para el EMLT dada en la ecuación (3.2). Esta expresión define EMLT como el cociente entre los parámetros de forma reducida y de primera etapa, es decir:

$$\lambda = \frac{\rho}{\phi},$$

lo que implica a su vez que la forma reducida, $\rho$, es igual al producto de la primera etapa por el EMLT:

$$\rho = \phi\lambda.$$

Se concluye así que en las muestras donde la primera etapa, $\phi$, es cero, la forma reducida debería ser cero también. Por otro lado, si se estima un valor no nulo de la forma reducida sin que haya indicios de un valor no nulo para la correspondiente primera etapa deberíamos preocuparnos, porque esto indica que hay algún otro canal, aparte de la variable de tratamiento (en este caso, el tamaño familiar) que enlaza los instrumentos con los resultados. Siguiendo esta idea, ALS identificó grupos demográficos para los que el efecto de las variables instrumentales «parto múltiple» e «igualdad de sexos» sobre el tamaño familiar fuera pequeño, de manera que no difiriera de cero de un modo significativo. Estas «muestras sin primera etapa» no generan indicios de que haya efectos significativos sobre las formas reducidas que pudieran apuntar a violaciones de la restricción de exclusión.

### *Los grandes almacenes de los mínimos cuadrados en dos etapas*

La esencia de la estimación de efectos causales con variables instrumentales consiste en efectuar comparaciones de la forma reducida entre grupos definidos por el instrumento, con la escala determinada por la correspondiente primera etapa. Este es un principio universal de las VI, pero los detalles pueden variar de una aplicación a otra. El estudio de la disyuntiva entre cantidad y calidad difiere del asunto KIPP en que disponemos de más de un instrumento para la misma

relación causal subyacente. Si se admite que los dos instrumentos «parto múltiple» e «igualdad de sexos» cumplen las condiciones necesarias y recogen de manera similar los efectos causales medios, nos gustaría combinar las dos estimaciones VI que generan para mejorar la precisión estadística. A la vez, el nacimiento de mellizos o gemelos podría estar correlacionado con ciertas características maternas, como la edad en el momento de dar a luz o el origen étnico, lo que sesgaría las estimaciones VI basadas en el instrumento «parto múltiple». Por eso nos gustaría disponer de algún procedimiento VI sencillo que controlara la edad de la madre y cualesquiera otros factores que pudieran inducir a error. Esto indica que pueden lograrse mejoras si se integran en el esquema VI los métodos de regresión tratados en el capítulo 2.

Los *mínimos cuadrados en dos etapas* (MC2E) generalizan los métodos basados en VI de dos maneras. Primero, los MC2E usan de manera eficaz múltiples instrumentos. Segundo, las estimaciones MC2E permiten controlar las variables explicativas, lo que mitiga el SVO de los instrumentos imperfectos. Para ver cómo funcionan los MC2E es útil reescribir los parámetros primera etapa *($\phi$)* y forma reducida *($\rho$)* como coeficientes de regresión, y no como diferencias de medias. Partamos de un solo instrumento, por ejemplo, una variable binaria $Z_i$ que indique si el segundo parto de una mujer ha sido múltiple. Entonces el efecto de la forma reducida puede escribirse como el coeficiente $\rho$ en una ecuación de regresión:

$$Y_i = \alpha_0 + \rho Z_i + e_{0i}. \tag{3.4}$$

Como se indica en el apéndice del capítulo 2, la regresión sobre un término constante y una única variable binaria calcula la diferencia entre las medias condicionadas de la variable dependiente cuando la variable binaria se activa o se desactiva. El coeficiente de $Z_i$ en la ecuación (3.4) es, por tanto,

$$\rho = E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0].$$

Del mismo modo, el efecto de la primera etapa de $Z_i$ es el coeficiente $\rho$ en la ecuación de la primera etapa:

$$D_i = \alpha_1 + \phi Z_i + e_{1i}, \tag{3.5}$$

donde $\phi = E[D_i|Z_i = 1] - E[D_i|Z_i = 0]$. Como $\lambda = \rho/\phi$, concluimos que el ELMT es el cociente de las pendientes de las regresiones (3.4) y (3.5).

El método MC2E brinda una alternativa para calcular $\rho/\phi$. El nombre MC2E procede del hecho de que el ELMT se puede obtener de una sucesión de dos regresiones. En la primera etapa se estima la ecuación (3.5) y se reservan los valores estimados, $\hat{D}_i$. Estas «estimaciones de la primera etapa» se definen como

$$\hat{D}_i = \sigma_1 + \phi Z_i. \tag{3.6}$$

La segunda etapa de MC2E plantea una regresión de $Y_i$ sobre $\hat{D}_i$ de este modo:

$$Y_i = \alpha_2 + \lambda_{\text{MC2E}}\hat{D}_i + e_2 i.$$

El valor de $\lambda_{\text{MC2E}}$ obtenido en este segundo paso es idéntico al cociente de los coeficientes correspondientes a la forma reducida y a la primera etapa, $\rho/\phi$, una relación teórica que se deduce en el apéndice de este capítulo.

En este procedimiento de regresión en dos pasos se pueden introducir con toda limpieza variables de control como la edad de la madre.[22] Si se añade la edad de la madre, representada como $A_i$, entonces la forma reducida y la primera etapa quedan así:

$$\text{Forma reducida: } Y_i = \alpha_0 + \rho Z_i + \gamma_0 A_i + e_{0i} \tag{3.7}$$
$$\text{Primera etapa: } \quad D_i = \alpha_1 + \phi Z_i + \gamma_1 A_i + e_{1i} \tag{3.8}$$

Aquí los ajustes de la primera etapa proceden de modelos que incluyen la variable de control $A_i$:

$$\hat{D}_i = \alpha_1 + \phi Z_i + \gamma_1 A_i.$$

[22] Ya hemos visto una versión de VI con covariables. Los efectos de la oferta de plazas en escuelas KIPP que constan en la columna (3) de la tabla 3.1 proceden de modelos de regresión para la primera fase y para la forma reducida que incluyen covariables en forma de variables binarias para distintos grupos de riesgo en las solicitudes.

Las estimaciones MC2E se construyen de nuevo planteando una regresión de $Y_i$ sobre las dos variables, $\{D \text{ circunflejo}\}_i$ y $A_i$. Por tanto, la ecuación de la segunda etapa MC2E es:

$$Yi = \alpha_2 + \lambda_{\text{MC2E}}\hat{D}_i + \gamma_2 A_i + e_{2i}, \tag{3.9}$$

que también incluye $A_i$.

El sistema MC2E permite incorporar tantas variables de control como se desee, siempre que aparezcan tanto en la primera etapa como en la segunda. Como se comenta en el apéndice del capítulo, el correspondiente ELMT se puede seguir construyendo a partir del cociente entre los coeficientes de forma reducida y de primera etapa, $\rho/\phi$. De hecho, deberíamos inspeccionar por separado el numerador y el denominador de este cociente para asegurarnos de que los dos se comportan como Dios manda. Pero cuando se trata de comunicar los resultados al público, el sistema MC2E es el método preferible incluso en problemas relativamente simples como este. Los paquetes informáticos de econometría calculan las estimaciones MC2E directamente, lo que reduce la posibilidad de cometer errores y permite estimar los errores típicos correspondientes sin esfuerzo adicional.[23]

¿Qué decir de nuestro segundo instrumento para el tamaño familiar, la variable binaria que marca la igualdad de sexos entre hermanos? Llamémosla $W_i$ (donde $W_i = 1$ indica igualdad de sexos, es decir, dos varones o dos niñas, y $W_i = 0$ sexos distintos). También aquí se introducen variables de control como, en particular, el sexo del primogénito, que se codifica en la variable binaria $B_i$ que se activa si el primogénito es varón (por regla general nacen algunos más varones que mujeres, así que la probabilidad de tener una pareja del mismo

[23] Los lectores atentos habrán notado que la variable de tratamiento que estamos manejando, el tamaño de la familia, no es una variable binaria como en el caso de la matriculación en centros KIPP, sino un tratamiento ordenado que cuenta hijos. Cabría preguntarse si es correcto describir las estimaciones MC2F de los efectos de tales variables como un ELMT. Aunque los detalles difieran, todavía se puede afirmar que las estimaciones MC2F recogen los efectos causales medios sobre los cumplidores en este contexto. La extensión del ELMT a tratamientos ordenados se desarrolla en Joshua D. Angrist y Guido W. Imbens, «Two Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity», *Journal of the American Statistical Association*, vol. 90, número 430, junio de 1995, páginas 431-442. De un modo similar, MC2F puede acomodar también variables instrumentales que no sean binarias. Veremos un ejemplo al respecto en el capítulo 6.

sexo es ligeramente superior cuando el primogénito es masculino). Con dos instrumentos, $W_i$ y $Z_i$, y la variable de control añadida $B_i$, la primera etapa se convierte en

$$D_i = \alpha_1 + \phi_t Z_i + \phi_s W_i + \gamma_1 A_i + B_i + e_{1i}. \tag{3.10}$$

Los efectos de la primera etapa de los instrumentos «parto múltiple» e «igualdad de sexos» se diferencian por los subíndices t para el primero y s para el segundo, y los escribimos como $\phi_t$ y $\phi_s$. Ambos instrumentos aparecen con los coeficientes marcados con tales subíndices en las formas reducidas correspondientes:

$$Y_i = \rho_1 Z_i + \rho_S W_i + \gamma_0 A_i + \delta_0 B_i + e_{0i}.$$

Con los ingredientes ya listos, ¡es hora de cocinar!

Las estimaciones de la segunda etapa con dos instrumentos y dos variables explicativas se generan con la ecuación de regresión

$$Yi = \alpha_2 + \lambda_{MC2E}\hat{D}_i + \gamma_2 A_i + \delta_2 B_i + e_{2i} \tag{3.11}$$

donde los valores estimados, $\hat{D}_i$, proceden de la ecuación de la primera etapa (3.10). Obsérvese que las variables explicativas aparecen siempre: en la primera y la segunda etapas y en la forma reducida. La ecuación (3.11) produce un promedio ponderado de las estimaciones que habríamos obtenido en caso de usar los instrumentos $Z_i$ y $W_i$ por separado, introduciendo las variables explicativas de control $A_i$ y $B_i$. Cuando los instrumentos generan resultados similares al usarlos por separado, entonces el promedio ponderado MC2E suele brindar una estimación más precisa de este efecto causal común.

El proceso MC2E ofrece un contexto maravillosamente flexible para la estimación por medio de variables instrumentales. Aparte de incorporar variables de control y de usar instrumentos múltiples con eficacia, este sistema permite el encaje de instrumentos de todos los tipos y tamaños, y no sólo variables binarias. Sin embargo, en la práctica, los maestros emplean paquetes informáticos estadísticos especialmente diseñados para calcular estimaciones MC2E, en lugar de efectuar regresiones sobre valores estimados como (3.11). La estimación de esta ecuación, conocida como «MC2E a mano», no produce

los errores típicos correctos que se requieren para medir la varianza muestral. El apéndice del capítulo explica por qué.

**Tabla 3.4. Primeras fases del estudio cantidad-calidad**

| | Instrumento parto múltiple | | Instrumento igualdad de sexos | | Ambos instrumentos juntos |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Segundo parto es múltiple | .320<br>(.052) | ,437<br>(.050) | | | .449<br>(.050) |
| Dos primeros hijos del mismo sexo | | | .079<br>(.012) | .073<br>(.010) | .076<br>(.010) |
| Varón | | –.018<br>(.010) | | –.020<br>(.010) | –.020<br>(.010) |
| Controles | No | Sí | No | Sí | Sí |

*Notas:* Esta tabla presenta los coeficientes de una regresión del número de hijos sobre instrumentos y variables explicativas. El tamaño de la muestra es 89.445. Los errores típicos constan entre paréntesis.

La tabla 3.4 muestra las estimaciones de las primeras etapas para los instrumentos «parto múltiple» e «igualdad de sexos», con y sin variables explicativas. La columna (2) presenta la estimación de un modelo de primera etapa con controles, y muestra que los primogénitos adultos israelíes cuyos hermanos siguientes fueron mellizos o gemelos se criaron en familias con unos 0,44 hijos más que los que crecieron en familias donde el segundo parto no fue múltiple. Esta estimación de primera etapa es superior a la estimación de 0,32 que se obtiene sin controles (y que consta en la columna [1]). La fórmula del SVO nos dice al respecto que el parto múltiple está asociado con factores que reducen el tamaño de la familia, como edades maternas mayores. Si se tiene en cuenta la edad de la madre, así como otros factores que pudieran causar confusión, se realza la primera etapa del instrumento «parto múltiple». Por otra parte, la primera etapa del instrumento «igualdad de sexos» procedente de un modelo con variables explicativas asciende a 0,073 y se parece al resultado sin controles, 0,079, porque el sexo de la descendencia carece de relación con los controles incorporados (estas estimaciones constan en las columnas [3] y [4]). El hecho de que el primogénito sea varón apenas repercute en el tamaño de la familia, tal como indica el pequeño valor,

sólo marginalmente significativo, de los coeficientes correspondientes que aparecen en la última fila de datos numéricos (se trata del único coeficiente de una variable de control cuyos valores se dan en la tabla, aunque la existencia de otros controles se indica en la fila inferior).[24]

Las estimaciones de segunda etapa de la disyuntiva cantidad-calidad constan en la tabla 3.5, junto con las estimaciones correspondientes dadas por una regresión convencional (es decir, sin instrumentos), por mínimos cuadrados ordinarios (MCO), de la forma

$$Y_i = \alpha_3 + \beta D_i + \gamma_3 A_i + \delta_3 B_i + e_{3i}.$$

Las estimaciones de la regresión convencional constan en la columna (1), y revelan una relación negativa importante entre el tamaño de la familia y el rendimiento educativo, incluso después de hacer ajustes con variables relacionadas con el origen étnico y la edad de la madre en el momento del nacimiento. En contraste, las estimaciones MC2F en este caso no resultan significativamente distintas de cero. La estimación con el instrumento «igualdad de sexos» refuerza los resultados obtenidos con el instrumento «parto múltiple». Las estimaciones MC2E que aparecen en la columna (3) muestran efectos generales positivos del tamaño familiar sobre la formación académica (aunque sólo una de las cantidades difiere de cero de manera significativa).

Un rasgo importante de ambas segundas etapas, tanto para el instrumento «igualdad de sexos» como para «parto múltiple», es su precisión, incluso a la hora de no detectar efectos. Los métodos VI descartan cualquier variación de la fertilidad, salvo aquella generada por el instrumento. Esto puede hacer que no haya suficiente variación para poder hallar resultados estadísticamente concluyentes. Pero aumentaremos la precisión si reunimos varios instrumentos, sobre todo si se da el caso de que esos instrumentos, tomados por separado, conducen a conclusiones similares (en este caso, tanto el instrumento

[24] Aparte de la variable binaria que indica si el primogénito es varón, entre las otras covariables se cuentan indicadores del año del censo, origen étnico de los progenitores, edad, si no consta el mes de nacimiento, edad de la madre, edad de la madre la primera vez que dio a luz, y edad de la madre en el momento de la inmigración (cuando esto es relevante). Véanse más detalles en el apartado de notas empíricas.

«parto múltiple» como «igualdad de sexos» arrojan pocos indicios en favor de la hipótesis de una disyuntiva cantidad-calidad). Las estimaciones agregadas de la primera etapa aparecen en la columna (5) de la tabla 3.4, mientras que los resultados correspondientes de la segunda etapa figuran en la columna (4) de la tabla (3.5).

**Tabla 3.5. Estimaciones del equilibrio cantidad-calidad mediante MCO y MC2E**

| | | Estimaciones MC2E | | |
|---|---|---|---|---|
| | Estimaciones MCO | Instrumento parto múltiple | Instrumento igualdad de sexos | Ambos instrumentos juntos |
| Variable dependiente | (1) | (2) | (3) | (4) |
| Años de escolaridad | –.145<br>(.005) | .174<br>(.166) | .318<br>(.210) | .237<br>(.128) |
| Graduado en secundaria | –.029<br>(.001) | .030<br>(.028) | .001<br>(.033) | .017<br>(.021) |
| Acudió a la universidad (para edad ≥ 24) | –.023<br>(.001) | .017<br>(.052) | .078<br>(.054) | .048<br>(.037) |
| Graduado universitario (para edad ≥ 24) | –.015<br>(.001) | –.021<br>(.045) | .125<br>(.053) | .052<br>(.032) |

*Notas:* Esta tabla presenta las estimaciones por MCO y por MC2E del efecto del tamaño familiar sobre la escolaridad. La columna (1) contiene las estimaciones por MCO. Las columnas (2), (3) y (4) reflejan las estimaciones MC2E construidas por medio de los instrumentos que se indican en las cabeceras de las columnas. Los tamaños muestrales son 89.445 para las filas (1) y (2), 50.561 para la fila (3) y 50.535 para la fila (4). Los errores típicos constan entre paréntesis.

Las estimaciones agregadas de la segunda etapa no son muy diferentes de las que se obtienen usando los instrumentos por separado, pero van acompañadas de errores típicos sensiblemente menores. Por ejemplo, el efecto estimado del tamaño familiar sobre el curso escolar más alto completado asciende a 0,24 cuando se usan ambos instrumentos, con un error típico de 0,13, una reducción considerable respecto de los errores típicos de 0,17 y 0,21 que resultan cuando se usan los instrumentos «parto múltiple» e «igualdad de sexos» por se-

parado. Es importante constatar que la estimación de la columna (1) para el curso escolar más alto realizado, un preciso –0,15, cae fuera del intervalo de confianza de la estimación MC2E que consta en la columna (4).[25] Esto parece indicar que la importante relación negativa observada entre el tamaño familiar y la formación académica se debe en buena medida, si no por completo, a sesgos de selección.

Maestro Joshway: Pequeño Saltamontes, levanta el hogar de las variables instrumentales.

Pequeño Saltamontes: Los cimientos tienen tres capas: (i) la *primera etapa* requiere instrumentos que afecten a la vía causal de interés; (ii) *el supuesto de independencia* requiere instrumentos que sean tan buenos como si hubiesen sido asignados al azar; (iii) *la restricción de exclusión* establece que sólo hay un camino causal que enlace los instrumentos con los resultados.

Maestro Joshway: ¿Es posible poner a prueba estos supuestos?

Pequeño Saltamontes: Comprobamos la primera etapa buscando una relación fuerte entre los instrumentos y la vía causal propuesta; comprobamos la independencia observando el cambio de las variables explicativas cuando el instrumento se activa o desactiva, como en un experimento aleatorio.

Maestro Joshway: ¿Qué hay de la exclusión?

Pequeño Saltamontes: No es fácil verificar la restricción de exclusión. Sin embargo, en ocasiones se encuentra una muestra en la que la primera etapa es muy pequeña. La exclusión requiere que en tales muestras las estimaciones para la forma reducida sean también pequeñas, porque el hipotético canal causal está ausente.

Maestro Joshway: ¿Cómo se calculan las estimaciones de las VI?

Pequeño Saltamontes: Los paquetes informáticos calculan por nosotros las estimaciones por mínimos cuadrados en dos etapas. Esto nos permite añadir variables explicativas y usar más de un

[25] En concreto, el coeficiente estimado por regresión vale –0,145 y cae fuera del intervalo de confianza del cálculo multiinstrumental MC2F, que da un resultado de $0{,}237 \pm (2 \times 0{,}128) = [-0{,}02, 0{,}49]$. Hay casos en los que se dispone de demasiados instrumentos, sobre todo si poseen escaso poder explicativo en la primera fase. El apéndice del capítulo amplía esta cuestión.

instrumento a la vez. Pero siempre tenemos que fijarnos en las estimaciones de la primera etapa y de la forma reducida.

## Maestros de la econometría: El increíble equipo de los Wright

El economista Philip G. Wright inventó el método de las VI con la ayuda de su hijo Sewall, que era genetista. Philip solía escribir sobre los mercados agrícolas. En 1928 publicó *The Tariff on Animal and Vegetable Oils*.[26] El libro se dedica sobre todo a esclarecer si los aranceles exorbitantes impuestos a principios de la década de 1920 a los productos de granja beneficiaban a los productores nacionales. Un crítico de la obra comentó en 1929 que «Sea cual fuere el valor práctico del intrincado cálculo de la elasticidad de la oferta y la demanda aplicada en particular a la mantequilla en este capítulo, el tratamiento es de un gran valor teórico».[27]

En los mercados competitivos, el desplazamiento simultáneo de las curvas de oferta y demanda conduce al equilibrio de precios y cantidades. No está claro cuál es el camino que lleva desde los precios y las cantidades de equilibrio hasta las curvas subyacentes de oferta y demanda que los generan. El problema de calcular las elasticidades de oferta y demanda a partir de la relación observada entre precios y cantidades se denomina el *problema de la identificación*. La identificación econométrica era una cuestión poco estudiada en la época en que Philip escribió. Los economistas tan sólo sabían con seguridad que la relación observada entre precio y cantidad no permitía identificar ni la oferta, ni la demanda, aunque estuviera determinada de algún modo por ambas.

El apéndice B de *The Tariff on Animal and Vegetable Oils* comienza con un planteamiento elegante del problema de la identificación en modelos de ecuaciones simultáneas. El apéndice pasa luego a explicar de qué modo resuelven el problema de la identificación las variables presentes en una ecuación pero ausentes en la otra. Philip se refirió a esas variables ausentes como «factores externos», porque al

[26] Philip G. Wright, *The Tariff on Animal and Vegetable Oils*, Macmillan Company, 1928.

[27] G. O. Virtue, «The Tariff on Animal and Vegetable Oils by Philip G. Wright», *American Economic Review*, vol. 19, número 1, marzo de 1929, páginas 152-156. La cita procede de la página 155.

manipular la ecuación en la que aparecen van trazando la ecuación en la que no constan (es decir, aquella para la cual son «externas»). Hoy llamamos *instrumentos* a esos agentes trazadores. Philip derivó y usó variables instrumentales para estimar las curvas de oferta y demanda en los mercados de mantequilla y linaza (la linaza se usa para producir un aceite utilizado en la fabricación de pinturas). El análisis de Philip del mercado de la linaza recurre al precio de los sustitutos de ese producto como trazadores de la demanda, mientras que la producción agrícola por unidad de superficie traza la oferta, que viene determinada por las condiciones meteorológicas.

El apéndice B supuso un avance trascendental, destacado e inesperado, en el pensamiento econométrico, hasta el punto de que hay quien se pregunta si realmente fue Philip quien lo escribió. Quizá su autor fuera Sewall, un estudioso muy distinguido por derecho propio. Como los maestros de la econometría Galton y Fisher, cuyas semblanzas aparecen al final de los capítulos 1 y 2, Sewall era genetista y estadístico. Sewall había desarrollado un método estadístico llamado «análisis de camino» mucho antes de la publicación del apéndice B, con la intención de resolver problemas relacionados con el sesgo de variables omitidas. Hoy se reconoce el análisis de camino como una aplicación de los métodos de regresión múltiple descritos en el capítulo 2, pero no resuelve el problema de la identificación planteado por los modelos de ecuaciones simultáneas. Algunos pasajes del apéndice B mencionan la idea de Sewall de «coeficientes de camino», pero el método de Philip de los factores externos era totalmente nuevo.

Los maestros James Stock y Francesco Trebbi investigaron la posible autoría de Sewall por medio de la estilometría.[28] Esta disciplina identifica a los autores por medio de regularidades estadísticas en el uso de las palabras y la estructura de las oraciones. La estilometría confirma que el autor del apéndice B fue Philip. Pero hace poco que Stock y su alumno Kerry Clark descubrieron un intercambio epistolar entre padre e hijo que muestra que las ideas del apéndice B las desarrollaron de manera conjunta en un toma y daca con grandes dosis de humildad. Philip describe en este intercambio el poder y la

[28] James H. Stock y Francesco Trebbi, «Who Invented Instrumental Variables Regression?», *Journal of Economic Perspectives*, vol. 17, número 3, verano de 2003, páginas 177-194.

simplicidad de las variables instrumentales, pero se muestra pesimista acerca de la facilidad con que podría aplicarse el método. En una carta a Sewall de 1926 en la que escribe acerca de las posibilidades de encontrar factores externos, Philip comentaba: «Me temo que tales factores, sobre todo en el caso de las condiciones de demanda, no son fáciles de encontrar».[29] La búsqueda de la identificación no se ha tornado más sencilla en las décadas transcurridas desde entonces.

El viaje de Philip era tanto personal como intelectual. Trabajó durante muchos años como profesor en un centro poco conocido, Lombard College, en Galesburg, Illinois. Lombard College no sobrevivió a la Gran Depresión, pero el tiempo que Philip pasó allí rindió unos frutos impresionantes. En Lombard tuvo como discípulo al joven Carl Sandburg, cuya poesía evocadora y de estructura relajada lo convirtió en un admirado símbolo estadounidense. He aquí la descripción que hizo Sandburg del camino abierto por la experiencia:[30]

> Esta mañana consulté el mapa del día
> y me dije: «¡Este es el camino! Por aquí es por donde iré,
> así recorreré las sendas del éxito,
> el camino está claro, será un placer seguir las líneas marcadas».
> Y caminando llegué a un sitio extraño,
> ¡un lugar que no estaba en el mapa!
> Tropecé, me caí y quedé tumbado sobre la maleza,
> y me arrepentí de aquella jornada.
>
> Aprendo poco a poco, aunque nunca estoy seguro,
> que sólo hay certidumbre en lo que ya ha pasado,
> que hay que asomarse al porvenir de vez en cuando
> como un vagabundo que deambula en la noche
> bajo una maraña de estrellas que no envían señales,
> y que ningún camino es seguro.

[29] Esta cita y la que consta en la figura proceden de cartas inéditas descubiertas por James H. Stock y Kerry Clark. Véase «Philip Wright, the Identification Problem in Econometrics, and Its Solution», presentado en un acto especial en honor de Philip Green en el Departamento de Economía de la Universidad Tufts en octubre de 2011 (http://ase.tufts.edu/economics/documents/wrightPhilipAndSewall.pdf), y la tesis de Kerry Clark «The Invention and Reinvention of Instrumental Variables Regresion».

[30] «Experience». Del libro *In Reckless Ecstasy*, Asgard Press, 1904, edición y prólogo de Philip Green Wright.

Veo conversar a hombres con mapas
que te indican adónde ir, por dónde y por qué,
mis oídos escuchan lo que dicen sus labios
mientras resueltos señalan las marcas de los mapas.
Pero sólo al que parezca robusto, solitario y quejumbroso
como si hubiera visitado un lejano país
y trazó su propio mapa
le diré: «¡Tu mapa es el que quiero ver!,
¡Me dejaría guiar por el mapa que tú tienes!».

## Apéndice: Teoría de variables instrumentales

### *VI, ELMT y MC2E*

Refresquemos antes que nada la notación del procedimiento de VI con un instrumento y sin variables explicativas. La primera etapa enlaza el instrumento con el tratamiento:

$$D_i = \alpha_1 + \phi Z_i + e_{1i}.$$

La forma reducida relaciona el instrumento con los resultados:

$$Y_i = \alpha_0 + \rho Z_i + e_{0i}.$$

La segunda etapa de los MC2E consiste en una regresión de los resultados sobre los valores estimados de la primera etapa:

$$Y_i = \alpha_0 + \lambda \hat{D}_i + e_{2i}.$$

Obsérvese que la fórmula (3.2) del ELMT se puede escribir tratando la primera etapa y la forma reducida como coeficientes de una regresión:

$$\lambda = \frac{\rho}{\phi} = \frac{C(Y_i, Z_i)/V(Z_i)}{C(D_i, Z_i)/V(Z_i)} = \frac{C(Y_i, Z_i)}{C(D_i, Z_i)}. \tag{3.12}$$

Aquí hemos recurrido al hecho de que las diferencias de medias del numerador y el denominador de la ecuación (3.2) coinciden con los

coeficientes de regresión $\phi$ y $\rho$. Cuando se escribe de este modo, es decir, como un cociente de covarianzas, la expresión de $\lambda$ recibe el nombre de *fórmula VI*. Su análogo muestral sería el estimador VI.

En este caso simple, la regresión de $Y_i$ sobre $\hat{D}_i$ (la segunda etapa en los MC2E) es lo mismo que la ecuación (3.12). Esto resulta evidente cuando se escribe la segunda etapa MC2F:

$$\lambda_{\text{MC2E}} = \frac{C(Y_i, \hat{D}_i)}{V(\hat{D}_i)} = \frac{C(Y_i, \alpha_1 + \phi Z_i)}{V(\alpha_1 + \phi Z_i)}$$

$$= \frac{\phi C(Y_i, Z_i)}{\phi_2 V(Z_i)} = \frac{\rho}{\phi} = \lambda.$$

Para deducirlo se recurre a las reglas de las varianzas y las covarianzas explicadas en el apéndice del capítulo 2.

Cuando se incluyen variables explicativas en las dos etapas, como por ejemplo la variable $A_i$ en nuestra investigación de la bomba poblacional, la segunda etapa de los MC2E es la ecuación (3.9). También ahora los MC2E equivalen a la fórmula VI, y esta última sigue dando el cociente de los coeficientes de la forma reducida y de la primera etapa. En este caso los coeficientes se estiman incluyendo $A_i$, como en las ecuaciones (3.7) y (3.8):

$$\frac{\rho}{\phi} = \frac{C(Y_i, \tilde{Z}_i)/V(\tilde{Z}_i)}{C(D_i, \tilde{Z}_i)/V(\tilde{Z}_i)} = \lambda_{\text{MC2E}},$$

donde $\tilde{Z}_i$ es el residuo de una regresión de $Z_i$ sobre $A_i$ (esto lo sabemos por la anatomía de la regresión). Los detalles que explican al segundo signo de igualdad se dejan al lector como ejercicio.

### *Errores típicos en los MC2E*

Como en las medias muestrales y las estimaciones por regresión, esperamos que las estimaciones VI y mediante MC2E varíen de una muestra a otra. Debemos calibrar el alcance de la variabilidad muestral para cualquier conjunto concreto de estimaciones para decidir si son significativas. La varianza muestral en las estimaciones mediante MC2E se cuantifican por medio de los correspondientes errores típicos.

Los errores típicos MC2E para un modelo que use $Z_i$ como instrumento para $D_i$, a la vez que incluye $A_i$ como control, se calculan del modo siguiente. Primero se construyen los residuos de los MC2E por medio de

$$\eta_i = Y_i - \alpha_2 - \lambda_{\text{MC2E}} D_i - \gamma_2 A_i.$$

El error típico para $\hat{\lambda}_{\text{MC2E}}$ viene dado entonces como

$$SE(\hat{\lambda}_{\text{MC2E}}) = \frac{\sigma_\eta}{\sqrt{\eta}} \times \frac{1}{\sigma_{\hat{D}}}, \tag{3.13}$$

donde $\sigma_\eta$ es la desviación típica de $\eta_i$, y donde $\sigma_{\hat{D}}$ es la desviación típica de los valores estimados de la primera etapa, $\hat{D}_i = \alpha_i + \phi Z_i + \gamma_1 A_i$.

Es importante remarcar que $\eta_i$ no es el residuo generado por la estimación manual de la segunda etapa de los MC2E, ecuación (3.9). Este residuo incorrecto sería

$$e_{2i} = Y_i - \alpha_2 - \lambda_{\text{MC2E}} \hat{D} - \lambda_2 A_i.$$

La varianza de $e_{2i}$ no interviene en la ecuación (3.13), así que el cálculo a mano de la segunda etapa en los MC2E generaría errores típicos incorrectos. La moraleja está clara: explore con libertad en la intimidad de su computadora privada, pero a la hora de extraer las estimaciones y los errores típicos que se harán públicos, deje que los paquetes informáticos profesionales le hagan el trabajo.

### *Sesgos en los MC2E*

Las variables instrumentales son una herramienta poderosa y flexible, pero los maestros recurren a sus armas más eficaces con prudencia. Como hemos visto, los MC2E combinan múltiples instrumentos en un esfuerzo por generar estimaciones precisas de un único efecto causal. Lo normal es que un investigador que cuente con la bendición de disponer de varios instrumentos sepa que algunos conducen a primeras etapas más potentes que otras. Existe la tentación de usar todos los instrumentos de todos modos (los paquetes informáticos de

econometría no salen más caros por eso). Pero se corre el riesgo de que las estimaciones por MC2E con muchos instrumentos débiles conduzcan a resultados equivocados. Un instrumento débil es aquel que no presenta una correlación fuerte con el regresor para el cual se aplica, de manera que el coeficiente de primera etapa resulta pequeño, o queda estimado de un modo poco preciso. Las estimaciones por MC2E con muchos instrumentos de este estilo tienden a parecerse a las estimaciones por MCO del mismo modelo. Cuando los MC2E se parecen a los MCO es natural concluir que no hay por qué preocuparse por sesgos de selección en estos últimos, pero esta conclusión podría ser errónea. Debido al *sesgo de muestra finita*, las estimaciones por MC2E en un modelo con muchas variables instrumentales débiles dicen poco acerca de la relación causal de interés.

¿Cuándo hay que preocuparse por el sesgo de muestra finita? Los maestros suelen fijarse en el estimador estadístico $F$ de la primera etapa para poner a prueba la hipótesis conjunta de que todos los coeficientes de la primera etapa de un conjunto de variables instrumentales son cero (el estimador estadístico $F$ generaliza el estimador $t$ para probar hipótesis múltiples). Una regla general muy utilizada exige que $F$ alcance como mínimo el valor 10 para perder el miedo a una multiplicidad de variables instrumentales débiles. Una alternativa a los MC2E la ofrece el estimador de por máxima verosimilitud con información limitada (MVIL), que se ve menos afectado por el sesgo de muestra limitada. Nos gustaría que las estimaciones por MVIL y por MC2E se parecieran entre sí, porque es poco probable que los primeros estén sesgados aun cuando se usen muchos instrumentos débiles (aunque las estimaciones por MVIL suelen conllevar errores típicos más elevados que las correspondientes estimaciones por MC2E).

El problema de la pluralidad de instrumentos débiles pierde virulencia cuando se usa un solo instrumento para estimar un único efecto causal. Las estimaciones relacionadas con la disyuntiva cantidad-calidad que usan, o bien una sola variable binaria para los partos múltiples, o bien una sola variable binaria para la igualdad de sexos como instrumento para el tamaño familiar, es poco probable que se vean afectadas por el sesgo de muestra finita. Estas estimaciones aparecen en las columnas (2) y (3) de la tabla 3.5. Para terminar, las estimaciones de formas reducidas merecen siempre una consideración atenta, porque se trata de estimaciones procedentes de MCO, por tanto no

afectadas por sesgos de muestra finita. Las estimaciones de formas reducidas que arrojan valores pequeños y que no difieren de cero de manera significativa, proporcionan un indicio fuerte y no sesgado de que la relación causal de interés es asimismo débil o inexistente, por lo menos en los datos que se están manejando (también se pueden poner a prueba múltiples coeficientes de formas reducidas de manera conjunta por medio de la función *F*). Siempre decimos a nuestros alumnos: *«Si no lo ves en la forma reducida, entonces es que no está»*.

4

# Diseños de regresión discontinua

Joven Caine: Maestro, conversemos más a fondo sobre las fuerzas del destino.

Maestro Po: Habla.

Caine: Cuando nos vemos de pie ante dos caminos por delante, ¿cómo esclarecer si será el camino izquierdo o el derecho el que nos conducirá a nuestro destino?

Maestro Po: Tú hablas del azar, Pequeño Saltamontes. Como si estuviera claro que tal cosa existe. En el asunto al que te refieres, el destino, no existe el azar.

*Kung Fu*, temporada 3, episodio 62

## *Nuestro camino*

El comportamiento humano está sometido a reglas. El estado de California limita el número de alumnos por aula a 32; con 33 ya sobra uno. La administración de la Seguridad Social de Estados Unidos no abonará ni un céntimo a ninguna persona que se retire antes de cumplir 62 años. Los aspirantes a reclutas para las fuerzas armadas cuyas notas en los exámenes caigan en los deciles inferiores no son elegibles para el ejército de Estados Unidos. Aunque muchas de estas normas nos parezcan arbitrarias y mal fundamentadas en la ciencia o la experiencia, decimos: ¡Adelante con ellas! Porque las normas que restringen el papel del azar en los asuntos humanos suelen generar experimentos interesantes. Los maestros de la econometría explotan

estos experimentos con una herramienta llamada *diseños de regresión discontinua* (RD). La RD no funciona con todas las cuestiones causales, pero sí lo hace con muchas de ellas. Y cuando funciona, sus resultados tienen casi tanta fuerza causal como los procedentes de un experimento aleatorio.

## 4.1 Cumpleaños y funerales

Katy: ¿De verdad es esto lo que quieres hacer el resto de tu vida?

Boon: ¿A qué te refieres?

Katy: A esto de ir por ahí con una pandilla de bestiajos que se emborrachan cada fin de semana.

Boon: ¡No! Cuando me gradúe pienso emborracharme cada noche.

*Desmadre a la americana*, 1978… por supuesto.

Cumplir 21 supone un hito muy importante en Estados Unidos, porque a partir de esa edad se pueden consumir bebidas alcohólicas «por fin», como dirían algunos. Por supuesto, también se bebe por debajo de esa edad. Y como ponen de manifiesto las burradas de Boon y sus compañeros de colegio mayor, quienes aún no tienen la edad legal no siempre beben con moderación. Un grupo de rectores de universidades estadounidenses está presionando a los estados para que la edad mínima legal para el consumo de alcohol (MLDA[1] por sus siglas en inglés) vuelva a ser la anterior a la guerra de Vietnam, 18 años, con la intención de resolver los problemas sociales y de salud asociados con el consumo de alcohol antes de alcanzar la edad legal. Esta postura, conocida como Iniciativa Amatista, se basa en la teoría de que situar la edad legal en los 18 años resta atractivo a saltarse las normas sobre consumo de alcohol y promueve una cultura madura al respecto. Esto contrasta con la visión tradicional de que la prohibición hasta los 21, aunque constituya una herramienta roma e imperfecta, reduce el acceso de la juventud al alcohol y, por tanto, evita ciertos daños.

Por fortuna, la historia de la MLDA genera dos experimentos naturales que se pueden utilizar para un diagnóstico fabuloso de las

[1] MLDA: *minimum legal drinking age. (N. de la T.).*

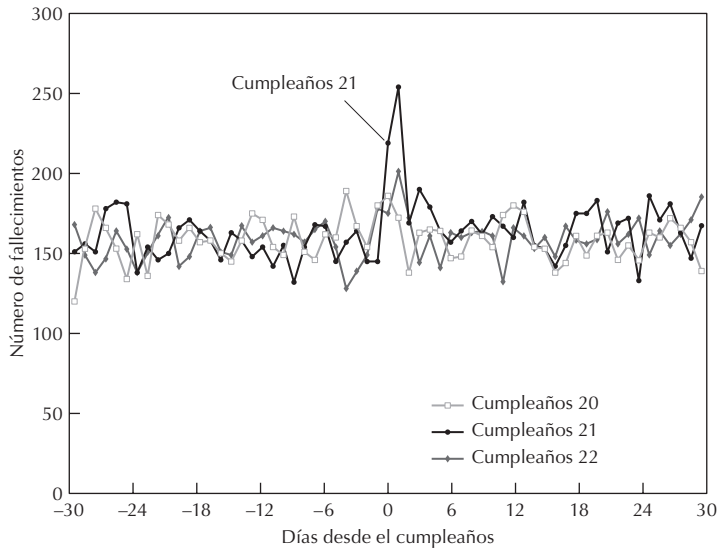



**Figura 4.1.** Cumpleaños y funerales.

políticas sobre consumo de alcohol. En este capítulo tratamos el primero de estos experimentos, mientras que el segundo se detalla en el capítulo siguiente.[2] El primer experimento MLDA surge del hecho de que un cambio pequeño en cuanto a edad (medido en meses, o incluso en días) genera un cambio enorme en cuanto a legalidad de acceso. La figura 4.1 muestra la diferencia inducida por un solo día, al representar la relación entre cumpleaños y funerales. Esta gráfica refleja el número de muertes de ciudadanos estadounidenses con edades comprendidas entre 20 y 22 años, desde 1997 hasta 2003. Se consignan las muertes para cada día, contando a partir de los cumpleaños, a los que se asigna el día 0. Por ejemplo, alguien que hubiera nacido el 18 de septiembre de 1990 y que muriera el 19 de septiembre de 2012 se computaría entre las muertes de personas de 22 años de edad sucedidas en el día 1.

[2] Nuestro tratamiento de la MLDA se basa en los trabajos de Christopher Carpenter y Carlos Dobkin, «The Effect of Alcohol Consumption on Mortality: Regression Discontinuity Evidence from the Minimum Drinking Age», *American Economic Journal – Applied Economics*, vol. 1, número 1, enero de 2009, páginas 164-182, y «The Minimum Legal Drinking Age and Public Health», *Journal of Economic Perspectives*, vol. 25, número 2, primavera de 2011, páginas 133-156.

El riesgo de muerte se dispara en los días inmediatamente posteriores al 21 cumpleaños, un hecho que se hace visible en el pico tan pronunciado del número de muertes diarias en esos días. Este pico añade unas 100 muertes a una línea de base de unas 150 por jornada. El pico de los 21 años no parece ser un efecto de excesos en las celebraciones de cumpleaños de carácter genérico, porque si este rasgo se debiera tan sólo a las fiestas de aniversario cabría esperar que las muertes se dispararan igualmente tras los cumpleaños 20 y 22, lo cual no ocurre. Sucede algo especial con el 21 cumpleaños, pero está por ver si este efecto puede atribuirse a la MLDA, y si el riesgo elevado de mortalidad que se aprecia en la figura 4.1 perdura lo bastante como para que valga la pena preocuparse por él.

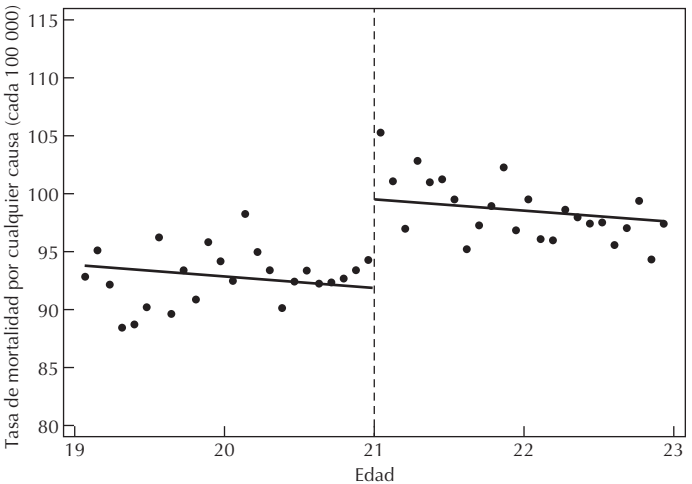



**Figura 4.2.** Estimación de los efectos de mortalidad de la MLDA mediante RD brusca.
*Notes:* This figure… = *Notas:* Esta figura representa la tasa de mortalidad por cualquier causa en función de la edad en meses. Las líneas de la gráfica representan los valores estimados mediante una regresión de la mortalidad incluyendo una variable binaria que marca si la edad en meses está o no por encima de los 21 años (la línea vertical discontinua indica el umbral de edad mínima legal para el consumo de alcohol, MLDA).

### *RD brusca*

La figura 4.2 aporta el argumento que vincula la MLDA con un incremento brusco y sostenido de la tasa de mortalidad. Esta gráfica representa las tasas de mortalidad (medidas en fallecimientos por cada 100 000 personas y por año) en función de la edad medida en meses (definidos como intervalos de 30 días), centrando la medida alrededor del 21 cumpleaños. El eje *X* abarca dos años a cada lado, y cada punto de la gráfica señala la mortalidad en un intervalo mensual. La mortalidad fluctúa de un mes a otro, pero hay pocos datos a la izquierda de la edad de 21 años que rebasen el 95. Sin embargo, por encima de esa edad la mortalidad se eleva, y hay pocos puntos a la derecha de los 21 años que bajen de 95.

Hay que alegrarse de que las probabilidades de que una persona joven fallezca descienden con la edad, un hecho que se aprecia en las pendientes descendentes de las rectas estimadas de las tasas de mortalidad representadas en la figura 4.2. Pero si extrapoláramos la línea de la izquierda, cabría esperar que a los 21 años de edad la mortalidad estuviera en torno a 92, mientras que la línea de la derecha parte de una cifra sensiblemente superior, alrededor de 99. El salto en las líneas de tendencia a los 21 años de edad ilustra el tema de este capítulo, los diseños de regresión discontinua (abreviado como diseños RD). La RD se basa en la idea, en apariencia paradójica, de que las normas rígidas (que a primera vista cabría pensar que reducen o incluso eliminan cualquier posible aleatoriedad) en realidad crean experimentos valiosos.

La cuestión causal planteada por la figura 4.2 radica en el efecto del acceso legal al alcohol sobre las tasas de mortalidad. La variable de tratamiento podría escribirse en este caso como $D_a$, donde $D_a = 1$ significa que se tiene la edad legal para beber, y $D_a = 0$ que aún no es así. $D_a$ es una función de la edad en años, $a$: la MLDA transforma a las personas de 21 años de menores de edad en consumidores de alcohol. Captamos esta transformación en notación matemática escribiendo:

$$D_a = \begin{cases} 1 & \text{si } a \geq 21 \\ 0 & \text{si } a < 21. \end{cases} \tag{4.1}$$

Esta representación realza dos rasgos característicos de la RD:

- El estado de tratamiento lo marca una función determinista de $a$, de manera que si se conoce $a$ se deduce $D_a$.
- El estado de tratamiento es una función discontinua de $a$, porque con independencia de lo cerca que esté $a$ del punto de corte, $D_a$ sigue sin cambiar hasta que se alcanza ese umbral.

La variable que determina el tratamiento, la edad en este caso, se denomina *variable móvil*. Las variables móviles desempeñan un papel central en el guión RD. En los diseños de RD *brusca* el tratamiento cambia con claridad de inactivo a activo cuando la variable móvil alcanza un cierto umbral. La MLDA es una función brusca de la edad, de manera que el estudio de los efectos de la MLDA sobre la mortalidad constituye un estudio de RD brusca. La segunda mitad de este capítulo trata un segundo guión RD, la *RD difusa*, en el que lo que cambia en el umbral es la probabilidad o la intensidad de tratamiento.

Está claro que la mortalidad cambia con la variable móvil, $a$, por razones que no están relacionadas con la MLDA. Las tasas de mortalidad debidas a enfermedades como el cáncer (conocidas en epidemiología como causas internas) son bajas, pero crecen para las personas de diecimuchos y veintipocos años, mientras que las muertes debidas a causas externas, sobre todo los accidentes de tráfico, homicidios y suicidios, descienden. Hay que separar esta tendencia al cambio de cualquier posible efecto de la MLDA y, para ello, el análisis RD incluye controles que dan cuenta de la variación suave que induce $a$ en las tasas de mortalidad.

Un análisis RD simple de los efectos causales estimados de la MLDA usaría una regresión como

$$\bar{M}_a = \alpha + \rho D_a + \gamma a + e_a, \tag{4.2}$$

donde $\bar{M}_a$ es la tasa de mortalidad en el mes $a$ (de nuevo definimos los meses como un intervalo de 30 días contado a partir del 21 cumpleaños). La ecuación (4.2) incluye la variable binaria de tratamiento, así como un control lineal sobre la edad en meses. Los valores estimados de la ecuación (4.2) generan las líneas trazadas en la figura 4.2. La pendiente negativa, representada como $\gamma$, refleja el suave declive de la

mortalidad entre la gente joven a medida que madura. El parámetro $\rho$ traza el salto en la mortalidad a la edad de 21 años. La regresión (4.2) arroja una estimación para $\rho$ igual a 7.7. Si se tiene en cuenta que la tasa de mortalidad promedio ronda 95, esta estimación indica un incremento sustancial del riesgo en el umbral MLDA.

¿Es esta una estimación creíble del efecto causal de la MLDA? ¿No habría que controlar otras cosas? La fórmula del SVO nos dice que la diferencia entre la estimación de $\rho$ en esta regresión simple y los resultados de cualquier regresión larga dependen de la correlación entre $D_a$ y las variables añadidas. Pero la ecuación (4.1) nos dice que $D_a$ depende tan sólo de $a$. Si se admite que el efecto de $a$ sobre la mortalidad queda bien representado por una función lineal, podemos tener la certeza de que la regresión corta no está afectada por SVO.

La ausencia de SVO en la ecuación (4.2) es un beneficio que procede de la información interna: aunque el tratamiento no se asigna al azar, conocemos su procedencia. En concreto, el tratamiento viene determinado por la variable móvil como implicación del vínculo determinista comentado más arriba. La cuestión de la causalidad se convierte, por tanto, en si la relación entre la variable móvil y los resultados queda o no bien reproducida por una regresión con un control de la edad.

Aunque la RD recurre a métodos de regresión para estimar efectos causales, es preferible contemplar los diseños RD como una herramienta distinta, con diferencias importantes respecto de los métodos de regresión comentados en el capítulo 2. En ese capítulo comparábamos los resultados del tratamiento y de los controles como valores particulares de las variables de control, con la esperanza de que, tras tener en cuenta los controles, el tratamiento quede asignado de un modo tan bueno como al azar. Pero aquí no hay ningún valor de la variable móvil en el que nos detengamos a considerar las observaciones de las variables de tratamiento y de control. ¡Vaya, Pequeño Saltamontes! A diferencia de las estrategias de emparejamiento y regresión del capítulo 2, basadas en comparaciones entre tratamiento y control condicionadas a los valores de las variables explicativas, la validez de la RD reside en nuestra intención de extrapolar sobre los valores de la variable móvil, al menos para valores en el entorno del umbral en el que se activa el tratamiento.

La naturaleza local de estas comparaciones de entorno se pone de manifiesto en la figura 4.2. El salto en las líneas de tendencia en el umbral MLDA compara de manera implícita las tasas de mortalidad para personas a un lado y a otro de su 21 cumpleaños, pero siempre cerca del mismo. Dicho de otro modo, el experimento imaginario consiste en cambiar el acceso de los jóvenes al alcohol, en un mundo en el que el alcohol está disponible de manera libre para los adultos. Los resultados de este experimento, a pesar de ser relevantes para los debates actuales acerca de las políticas sobre el alcohol, no tienen mucho que decir acerca de las consecuencias de otros cambios políticos más drásticos, como la prohibición total del alcohol.

### *Particularidades de la RD*

No hay garantía de que las herramientas RD produzcan estimaciones causales fiables. La figura 4.3 explica por qué es así. En la gráfica A, la relación entre la variable móvil *(X)* y el resultado *(Y)* es lineal, con un claro salto en $E[X|Y]$ en el valor umbral igual a un medio. La gráfica B es muy parecida, salvo por el hecho de que la relación promedio entre $X$ e $Y$ no es lineal. Aun así se aprecia muy bien el salto en $X = 0{,}5$. La gráfica C de la figura 4.3 pone de manifiesto el desafío al que se enfrentan los diseños RD. Aquí la figura muestra una tendencia no lineal más barroca, con giros bruscos a la izquierda y la derecha del umbral, pero sin discontinuidad. Las estimaciones que se deriven de un modelo lineal como el de la ecuación (4.2) confundirán la no linealidad con una discontinuidad.

Hay dos estrategias que reducen la probabilidad de cometer errores al aplicar RD, aunque ninguna proporciona una seguridad total. La primera modela directamente la no linealidad, mientras que la segunda se centra tan sólo en las observaciones cercanas al umbral. Empezaremos con la estrategia de modelos no lineales, y trataremos brevemente la otra hacia el final de este apartado.

Lo habitual en el contexto RD es tratar las no linealidades por medio de funciones polinómicas de la variable móvil. En el caso ideal, esta aproximación produce resultados insensibles al grado de no linealidad permitido por el modelo. Pero a veces, sin embargo, como en el caso de la gráfica C de la figura 4.3, no es esto lo que ocurre.

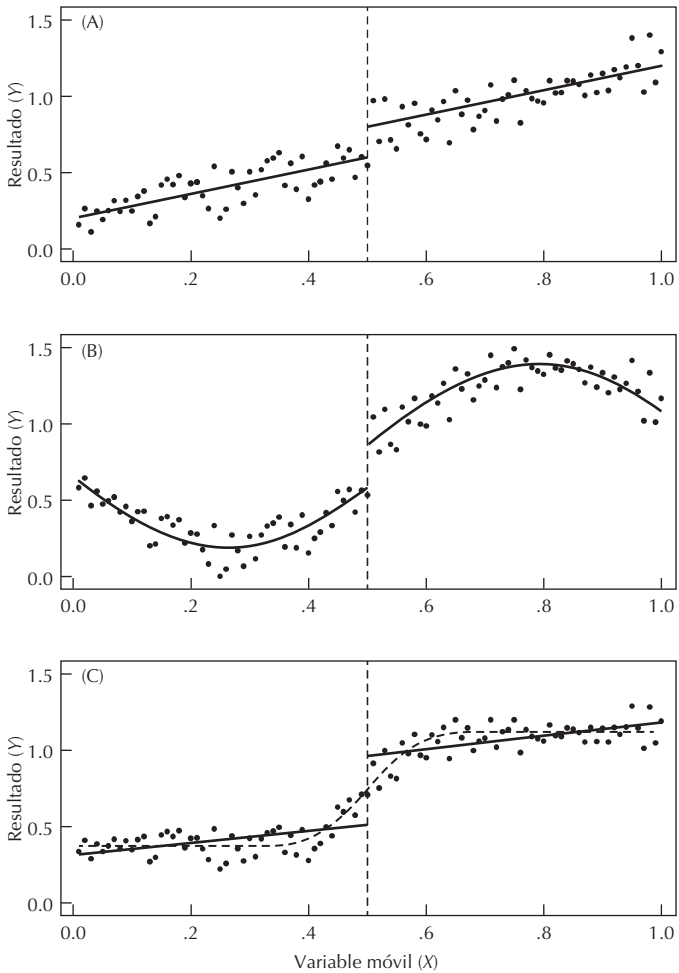



**Figura 4.3.** RD en acción, tres enfoques.
*Notas:* La gráfica A muestra una RD con un modelo lineal para $E[Y_i|X_i]$. La gráfica B añade una cierta curvatura. La gráfica C muestra una falta de linealidad malinterpretada como una discontinuidad. La línea vertical discontinua marca un umbral RD hipotético.

La cuestión de cuánta no linealidad es suficiente requiere un juicio sosegado. Se corre el riesgo de elegir el modelo que proporcione los resultados más llamativos, lo que quizá favorezca las alternativas que mejor cuadren con los prejuicios del investigador. Por tanto, quienes

aplican RD deben proporcionar a sus lectores información acerca del modo en que cambian las estimaciones RD a medida que se alteran los detalles del modelo de regresión empleado.

La figura 4.2 sugiere la posibilidad de que haya una curvatura suave en la relación entre $\bar{M}_a$ y $a$, al menos para los puntos que caen a la derecha del umbral. Una extensión simple que refleje esta curvatura podría recurrir a un control cuadrático, en lugar de a uno lineal, para la variable móvil. El modelo RD con control cuadrático de la variable móvil se convierte en

$$\bar{M}_a = \alpha + \rho D_a + \gamma_1 a + \gamma_2 a^2 + e_a,$$

donde $\gamma_1 a + \gamma_2 a^2$ es una función cuadrática de la edad, y las dos $\gamma$ son parámetros que hay que estimar.

Una modificación relacionada con la anterior permite diferentes coeficientes para la variable móvil a la izquierda y a la derecha del umbral. Esta modificación genera modelos de la interacción entre $a$ y $D_a$. Para facilitar la interpretación del modelo con interacción, se centra la variable móvil sustrayéndole el umbral, $a_0$. Al sustituir $a$ por $a - a_0$ (para nosotros $a_0 = 21$) y añadiendo un término de interacción, $(a - a_0)\, D_a$, el modelo RD queda:

$$\bar{M}_a = \alpha + \rho D_a + \gamma(a - a_0) + \delta[(a - a_0)D_a] + e_a. \qquad (4.3)$$

Al centrar la variable móvil nos aseguramos de que $\rho$ en la ecuación (4.3) siga representando el salto en resultados medios en el umbral (como se ve al hacer $a = a_0$ en la ecuación).

¿Por qué tendría que cambiar la tendencia de la relación entre edad y mortalidad en el umbral? Los datos a la izquierda del umbral reflejan la relación entre edad y mortalidad para una muestra cuyos hábitos de bebida están restringidos por la MLDA. En esta muestra cabría esperar un descenso paulatino de la mortalidad, a medida que la gente joven madura y tiende a asumir cada vez menos riesgos. Sin embargo, tras la edad de 21 años el acceso al alcohol sin restricciones podría alterar este proceso, quizá atenuando la tendencia a la bajada de la mortalidad. Por otra parte, si los rectores universitarios que apoyan la Iniciativa Amatista tuvieran razón, la responsabilidad de consumir alcohol legalmente aceleraría el desarrollo de comportamientos

más maduros. El sentido en que irían esos cambios en las pendientes no es más que una hipótesis, pero el punto clave está en que la ecuación (4.3) permite que esas pendientes cambien en cualquier sentido.

Una implicación sutil del modelo con términos de interacción es que al apartarse del umbral $a_0$ el tratamiento MLDA surte un efecto dado por $\rho + \delta(a - a_0)$. Esto se ve si se calcula la diferencia entre las rectas de regresión estimadas cuando $D_a$ no está activada y cuando sí lo está:

$$\begin{aligned}&[\alpha + \rho + (\gamma + \delta)(a - a_0)] - [\alpha + \gamma(a - a_0)]\\ &\quad = \rho + \delta(a - a_0).\end{aligned}$$

Sin embargo, las estimaciones lejos del umbral implican una extrapolación exagerada y deben tomarse con una rodaja de lima y un puñadito de sal. No disponemos de datos sobre tasas de mortalidad contrafactuales procedentes de un mundo en el que el consumo de alcohol quedara prohibido para edades superiores a 21 años. Del mismo modo, si nos apartamos del umbral hacia la izquierda costaría decir qué tasas de mortalidad cabría esperar en un mundo donde se permitiera beber desde edades muy tempranas. En contraste, parece razonable afirmar que las personas justo por debajo de la edad de 21 años proporcionan una buena comparación contrafactual frente a quienes están justo por encima de 21 años. Esto nos lleva a observar las estimaciones del parámetro $\rho$ (el efecto causal a la derecha del umbral) como las más fiables, incluso aunque el modelo empleado para la estimación nos diga mucho más que eso de manera implícita.

Las tendencias no lineales y los cambios de pendiente en el entorno del umbral se pueden combinar en un modelo que tenga este aspecto:

$$\begin{aligned}\bar{M}_a = \alpha + \rho\, D_a + \gamma_1(a - a_0) + \gamma_2(a - a_0)^2. \\ + \delta_1[(a - a_0)D_a] + \delta_2[(a - a_0)^2 D_a] + e_a.\end{aligned} \tag{4.4}$$

En este contexto tanto los términos lineales como los cuadráticos cambian al cruzar el umbral. Como antes, el salto en cuanto a tasas de mortalidad en el umbral MLDA se refleja en el efecto del tratamiento MLDA, $\rho$. El efecto del tratamiento al alejarse del umbral es ahora $\rho + \delta_1(a - a_0) + \delta_2(a - a_0)^2$, aunque de nuevo la interpretación

causal de esta cantidad es más especulativa que la interpretación causal de $\rho$ en sí misma.

La figura 4.4 muestra que la función de tendencia estimada que genera la ecuación (4.4) posee cierta curvatura, ligeramente cóncava a la izquierda de la edad de 21 años y marcadamente convexa después. Este modelo genera una estimación mayor que el lineal del efecto MLDA en el umbral, y que ahora asciende a 9,5 muertes cada 100.000. La figura 4.4 revela también la tendencia lineal generada por la ecuación (4.2). El modelo más elaborado parece ofrecer un ajuste mejor que el simple: la mortalidad asciende de manera brusca a la edad de 21 años, pero luego se recupera en cierta medida en los primeros meses tras el 21 aniversario. Esto recuerda al pico en las tasas de mortalidad diaria en el 21 cumpleaños o su entorno que se aprecia en la figura 4.1. A diferencia de Boon y sus compañeros de fraternidad, parece que muchos bebedores recién legalizados se acaban cansando de acostarse bebidos cada noche. La especificación (4.4) capta bien este salto y su declive, aunque al precio de volverse algo recargado desde el punto de vista técnico.

¿Qué modelo es mejor, el sofisticado o el simple? No hay reglas generales al respecto, ni nada que sustituya a la observación reflexiva de los datos. Hay que considerarse especialmente afortunados cuando los resultados no son especialmente sensibles a los detalles del modelo elegido, como parece suceder en la figura 4.4. El modelo RD simple parece lo bastante flexible como para captar los efectos justo en el umbral, en este caso en torno al 21 cumpleaños. La versión más compleja estima el pico en tasa de mortalidad en torno al cumpleaños número 21, y también registra la recuperación parcial posterior de esas tasas.

Los efectos en el umbral no son necesariamente lo más importante. Supongamos que se eleva la edad para beber hasta los 22 años. En un mundo donde las muertes por exceso de alcohol se debieran exclusivamente a las fiestas de cumpleaños MLDA, un cambio así alargaría algunas vidas en un año, pero no tendría ningún otro efecto. El incremento sostenido en las tasas de mortalidad que se aprecia en la figura 4.4 es, por lo tanto, importante, porque sugiere que restringir el acceso al alcohol induce beneficios duraderos. Ya comentamos antes que los indicios acerca de efectos alejados del umbral son más especulativos que los que se refieren al salto o a su entorno. Por otra

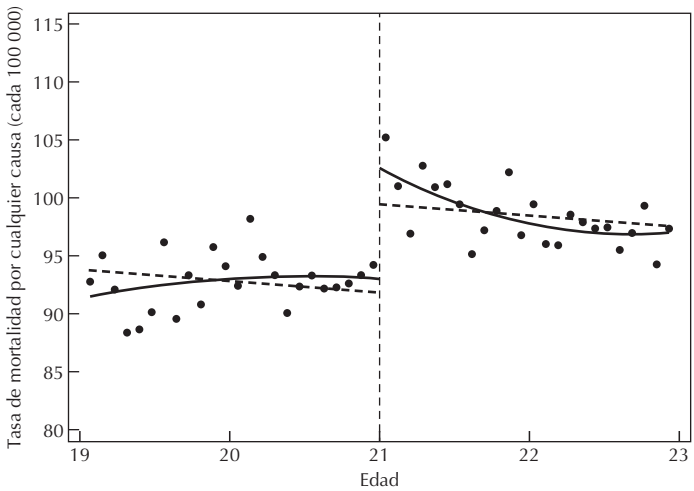



**Figura 4.4.** Control cuadrático en un modelo de RD.
*Notas:* Esta figura representa la tasa de mortalidad por cualquier causa en función de la edad en meses. Se emplean líneas gruesas discontinuas para representar los valores estimados con un modelo de regresión de la mortalidad sobre la edad en meses que incluye una variable binaria que marca si la edad es superior o no a 21 años. Las líneas continuas representan los valores estimados mediante una regresión de la mortalidad que incluye la misma variable binaria más un término cuadrático de la edad (la línea discontinua vertical marca el umbral de edad mínima legal para el consumo de alcohol, MLDA).

parte, cuando la relación entre la variable móvil y los resultados es aproximadamente lineal, una extrapolación limitada parece justificada. El salto en tasas de mortalidad en el umbral muestra que los hábitos de bebida responden al acceso al alcohol de un modo que queda reflejado en la mortalidad, un punto en principio importante, mientras que el efecto del tratamiento MLDA extrapolado hasta la edad de 23 años aún parece sustancial y creíble, del orden de cinco muertes anuales adicionales por cada 100.000 personas. Este patrón realza el valor de la «RD visual», es decir, una interpretación cuidadosa de los gráficos como el de la figura 4.4.

¿Hasta qué punto es convincente el argumento de que el salto de la figura 4.4 se debe de verdad a la bebida? Los datos que relacionan las tasas de mortalidad con las causas de fallecimiento nos ayudan a esclarecerlo. Aunque el alcohol sea nocivo, hay pocas personas que

mueran por mera intoxicación etílica, mientras que las muertes por patologías asociadas al alcohol se producen tan sólo a edades avanzadas. Pero el alcohol mantiene una relación estrecha con los accidentes de tráfico con vehículos motorizados (AVM), la primera causa de muerte en la gente joven. Si conducir ebrio fuera la causa primaria de muerte vinculada al consumo de alcohol, entonces deberíamos apreciar un gran incremento en las muertes por AVM junto a un cambio escaso en la mortalidad debida a causas internas. Como en las pruebas de equilibrio para el experimento RAND HIE de la tabla 1.3 y en el caso del instrumento para la oferta de plazas KIPP del apartado A de la tabla 3.1, un efecto nulo en los resultados que deberían permanecer inalterados por el tratamiento incrementará nuestra confianza en los efectos causales que estamos persiguiendo.

La tabla 4.1 muestra el banco de pruebas para los resultados relacionados con las distintas causas de muerte. La primera fila presenta estimaciones del total de fallecimientos construidas usando tanto la ecuación RD simple (4.2) como la compleja (4.4). Los resultados constan en las columnas (1) y (2). La segunda fila de la tabla 4.1 revela efectos fuertes del consumo legal de alcohol sobre las muertes por AVM, efectos lo bastante grandes como para dar cuenta del total del exceso de muertes relacionadas con la MLDA. Estas estimaciones son bastante insensibles al hecho de deducirlas con el modelo simple o con el complejo. Entre las otras causas de muerte que se podrían esperar relacionadas con el consumo de alcohol están el suicidio y otras causas externas, lo que incluye accidentes distintos a los de tráfico. De hecho, los efectos estimados sobre suicidios y muertes por otras causas externas (excluyendo los homicidios) también presentan algunos incrementos pequeños, pero estadísticamente significativos, en el umbral MLDA.

Es importante constatar que las estimaciones que constan en las columnas (1) y (2) sobre muertes por causas internas (lo que incluye muertes por cáncer y otras enfermedades) son pequeñas y no difieren significativamente de cero. Tal como muestra la última fila de la tabla, los efectos de la intoxicación etílica directa también parecen modestos y de una magnitud similar a la de las causas internas, aunque el salto estimado en las muertes por intoxicación etílica difiera de cero de un modo estadísticamente significativo. En conjunto, por tanto, la tabla 4.1 apoya la hipótesis MLDA al mostrar efectos claros

**Tabla 4.1. Estimaciones de los efectos de la MLDA sobre la mortalidad con RD brusca**

| Variable dependiente | Edades 19-22 | | Edades 20-21 | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Todas las muertes | 7.66 | 9.55 | 9.75 | 9.61 |
| | (1.51) | (1.83) | (2.06) | (2.29) |
| Accidentes de tráfico | 4.53 | 4.66 | 4.76 | 5.89 |
| | (.72) | (1.09) | (1.08) | (1.33) |
| Suicidio | 1.79 | 1.81 | 1.72 | 1.30 |
| | (.50) | (.78) | (.73) | (1.14) |
| Homicidio | .10 | .20 | .16 | –.45 |
| | (.45) | (.50) | (.59) | (.93) |
| Otras causas externas | .84 | 1.80 | 1.41 | 1.63 |
| | (.42) | (.56) | (.59) | (.75) |
| Todas las causas internas | .39 | 1.07 | 1.69 | 1.25 |
| | (.54) | (.80) | (.74) | (1.01) |
| Causas relacionadas con el alcohol | .44 | .80 | .74 | 1.03 |
| | (.21) | (.32) | (.33) | (.41) |
| Controles | edad | edad y edad$^2$ junto a variable binaria >21 | edad | edad y edad$^2$ junto a variable binaria >21 |
| Tamaño de la muestra | 48 | 48 | 24 | 24 |

*Notas:* Esta tabla recoge los coeficientes de regresiones de las tasas de mortalidad específicas por edades en meses y según su causa, sobre una variable binaria que marca edades superiores a 21 años, y añadiendo a la variable binaria controles de edad, o bien lineales, o bien cuadráticos. Los errores típicos constan entre paréntesis.

sobre las causas que parece más razonable atribuir al alcohol, pero pocos indicios de incrementos debidos a causas internas.

La figura 4.5 brinda más apoyo a estas conclusiones en una gráfica que representa el número de víctimas de AVM, construida usando el modelo que genera las estimaciones de la columna (2) de la tabla 4.1. La figura muestra una discontinuidad clara en el umbral MLDA, y sin indicios de posibles efectos no lineales que pudieran confundirnos. Al mismo tiempo, no se ve ningún salto en las muertes por causas internas, mientras que los errores típicos de la tabla 4.1 sugieren que lo más probable es que el ligero escalón que se aprecia en la figura para este tipo de fallecimientos se deba al azar.

Además de la estimación directa por regresión, hay otra aproximación que los maestros llaman RD paramétrica, una segunda estrategia de RD que explota el hecho de que el problema de distinguir entre saltos y tendencias no lineales se vuelve menos acuciante cuando ponemos el objetivo sobre puntos cercanos al umbral. No tenemos que preocuparnos en absoluto por tendencias no lineales para el pequeño conjunto de puntos cercanos a la frontera. Esto sugiere una aproximación que compara promedios dentro de ventanas estrechas justo a la izquierda y justo a la derecha del umbral. Aquí surge el inconveniente de que si la ventana es demasiado pequeña, entonces caen pocas observaciones en su interior, lo que torna probable que las estimaciones extraídas sean demasiado imprecisas para resultar útiles. Pero, aun así, deberíamos ser capaces de equilibrar la reducción del sesgo cerca de la frontera frente al incremento de varianza causado por el rechazo de datos, lo que conducirá a una medida óptima de algún tipo para el tamaño de la ventana.

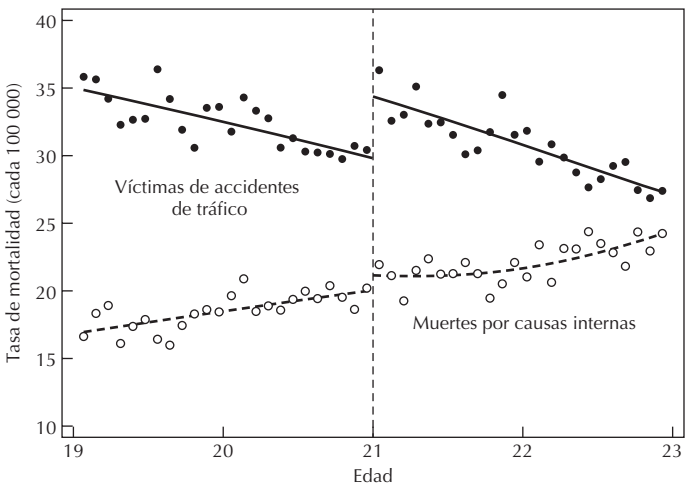



**Figura 4.5.** Estimación mediante RD del efecto de la MLDA sobre la mortalidad según causa de muerte.
*Notas:* Esta figura representa la tasa de mortalidad por accidentes de tráfico y por causas internas, en función de la edad en meses. Las líneas representan los valores estimados por regresión de la mortalidad, según su causa, sobre una variable binaria que marca si la edad supera los 21 años más una función cuadrática de la edad en meses (la línea discontinua vertical marca el umbral de edad mínima legal para el consumo de alcohol, MLDA).

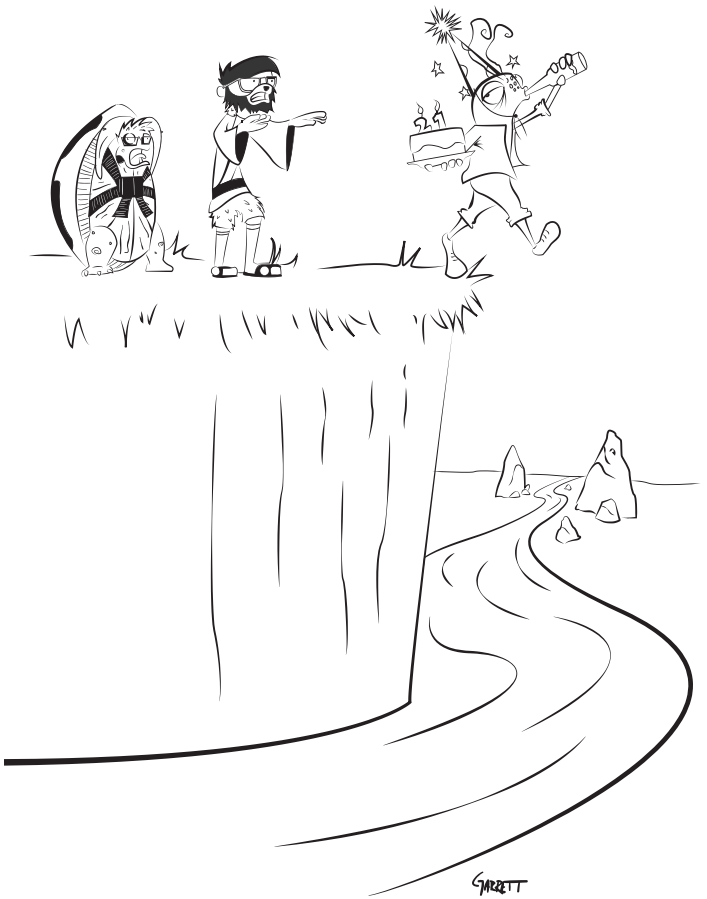
El proceso econométrico que valora este equilibrio es el de la *RD no paramétrica*. La RD no paramétrica consiste en estimar la ecuación (4.2) en una ventana estrecha en torno al umbral; es decir, estimamos:

$$\bar{M}_a = \alpha + \rho D_a + \gamma a + e_a; \quad \text{en una muestra tal que } a_0 - b \leq a \leq a_0 + b. \tag{4.5}$$

El parámetro *b* describe el tamaño de la ventana y se denomina *ancho de banda*. Cabe contemplar los resultados de la tabla 4.1 como una RD no paramétrica con un ancho de banda igual a dos años de edad para las estimaciones que constan en las columnas (1) y (2), y un ancho de banda la mitad de grande (es decir, que incluya sólo las edades de 20 a 21 años, en lugar de 19 a 22) para las estimaciones de las columnas (3) y (4). La elección del modelo simple en la ecuación (4.5) frente al complejo de la ecuación (4.4) debería importar poco cuando ambos se aplican a ventanas de edad estrechas en el entorno del umbral. Los resultados de la tabla 4.1 respaldan esta conjetura, aunque hay cierta inestabilidad en las estimaciones de una columna a otra que parece razonable atribuir a la varianza muestral.[3]

¡Es bien simple! Pero ¿cómo elegimos el ancho de banda? Por una parte, para disipar la preocupación por la elección del polinomio nos gustaría trabajar con los datos cercanos al umbral. Pero por otra parte, menos datos significa menos precisión. Por lo tanto, y para empezar, el ancho de banda debería variar en función del tamaño muestral. Cuanta más información tengamos acerca de los resultados en el entorno del umbral de la RD, más estrecho puede fijarse el ancho de banda manteniendo aún la esperanza de generar estimaciones lo bastante precisas como para que resulten útiles. La econometría teórica propone estrategias sofisticadas para hallar de manera eficaz este equilibrio entre sesgo y varianza, aunque tampoco en esos casos el algoritmo de selección del ancho de banda depende por completo

[3] Los gurús de la RD no paramétrica suelen estimar los modelos como el de la ecuación (4.2) usando mínimos cuadrados ponderados. Este procedimiento atribuye más peso a las observaciones más cercanas al umbral, y menos a las que caen más apartadas. La función de pesos que se usa para este fin se denomina *núcleo* (o *kernel*). Las estimaciones de la tabla 4.1 usan de manera implícita un *núcleo uniforme*, es decir, atribuyen el mismo peso a todas las observaciones que caen dentro del ancho de banda.




de los datos y hay que elegir ciertos parámetros.[4] En la práctica, la elección del ancho de banda (así como la del polinomio en los modelos paramétricos) precisa sentido común. El objetivo se centra no tanto en hallar el ancho de banda perfecto como en mostrar que las conclusiones generadas por cualquier elección concreta de ancho de banda no son resultado del azar.

En este sentido, los estudios en los que se basa nuestra investigación de la MLDA parecen elaborados en el paraíso de la regresión

[4] Véase Guido W. Imbens y Karthik Kalyanaraman, «Optimal Bandwidth Choice for the Regression Discontinuity Estimator», *Review of Economic Studies*, vol. 79, número 3, julio de 2012, páginas 933-959.

discontinua (quizá como recompensa por la templanza de los autores). Las estimaciones RD generadas por los modelos paramétricos con distintos polinomios de control resultan similares entre sí, y están cerca del conjunto correspondiente de estimaciones no paramétricas. Estas estimaciones no paramétricas son en su mayor parte insensibles a la elección del ancho de banda sobre un intervalo amplio.[5] Esta coincidencia de resultados sugiere que los hallazgos generados por el análisis de RD de la MLDA reflejan efectos causales reales. Algunas personas jóvenes parecen pagar el precio más alto posible por el privilegio de tomarse un trago legal.

### 4.2 La ilusión de la élite

Kwai Chang Caine: No aspiro a encontrar las respuestas, sino a comprender las preguntas.

*Kung Fu*, temporada 1, episodio 14

Los sistemas educativos públicos de Boston y de la ciudad de Nueva York incluyen un puñado de centros de acceso selectivo previo examen. A diferencia de muchos otros colegios públicos estadounidenses, los centros selectivos (*exam schools*) filtran las solicitudes sobre la base de una prueba de admisión de carácter competitivo. Del mismo modo que muchos graduados de la escuela secundaria compiten por matricularse en las universidades más selectivas del país, en algunas ciudades los estudiantes más jóvenes y sus progenitores aspiran a unas disputadas plazas en centros selectivos de élite. Menos de la mitad de quienes solicitan el acceso en la escuela selectiva de Boston logran una plaza en la John D. O'Bryant School, la Boston Latin Academy o la Boston Latin School (BLS); sólo una sexta parte de las solicitudes de Nueva York logran plaza en alguno de los

[5] Se puede consultar una comparación de las estimaciones paramétricas y no paramétricas en las tablas 4 y 5 de Carpenter y Dobkin, «The Effect of Alcohol Consumption», *American Economic Journal*: Applied Economics, 2009. La sensibilidad a la elección del ancho de banda se investiga en un apéndice disponible en Internet (DOI: 10.1257/app.1.1.164). El estudio de 2009 analiza la mortalidad a través del día exacto de nacimiento, mientras que aquí trabajamos con datos mensuales.

tres centros selectivos originales de la Gran Manzana (Stuyvesant, Bronx Science, Brooklyn Tech).

Esta competencia interna por las plazas parece comprensible a primera vista. Muchos estudiantes de centros selectivos siguen más tarde carreras brillantes en ciencias, artes o política. Los estudiantes de estas escuelas se encuentran muy por delante de los de otros centros públicos, se mida como se mida. Es fácil ver por qué muchos padres darían un riñón (¡o incluso el hígado!) por colocar a sus hijos en esos colegios. Los economistas y otros científicos sociales también tienen interés por las consecuencias del tratamiento en los centros selectivos. Ante todo, porque estas escuelas reúnen a los estudiantes más capacitados. Sin duda, esto es algo bueno: los estudiantes brillantes aprenden tanto de sus pares como de sus profesores, o esto es lo que afirman ciertas instituciones selectivas como el MIT o la London School of Economics.

Es fácil defender las ventajas de los centros selectivos, pero también está claro que al menos parte de la diferencia en cuanto a resultados que se asocia a estas escuelas refleja las propias políticas selectivas

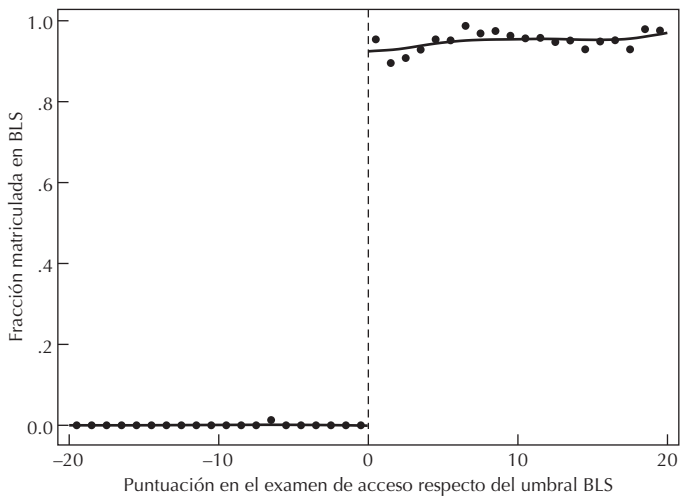



**Figura 4.6.** Matriculación en BLS.
*Notas:* Esta figura representa las tasas de matriculación en la Boston Latin School (BLS), en función de las notas en el examen de admisión, para quienes solicitaron plaza en BLS y obtuvieron calificaciones en el entorno del umbral. Las líneas continuas muestran valores estimados mediante regresión lineal local, con estimaciones separadas a cada lado del umbral (el cual se señala con la línea vertical discontinua).

de admisión. Si un centro admite sólo a gente excelente, entonces el alumnado que cursa allí los estudios será excelente, con independencia de si el colegio aporta o no valor añadido. Esto suena a sesgo de selección, y lo es. Siguiendo la estela de las amplias miras de las autoridades sanitarias de Oregón con su sorteo de seguros médicos, podríamos confiar en convencer a Stuyvesant y Boston Latin para que admitieran estudiantes al azar, y no basándose en un examen. Entonces podríamos utilizar los datos experimentales resultantes para conocer si los centros selectivos aportan valor añadido. ¿O no podríamos? Porque si los colegios selectivos admitieran estudiantes al azar, entonces dejarían de ser colegios selectivos después de todo.

Si el carácter selectivo del procedimiento de admisión forma parte necesaria de la esencia de un colegio selectivo, entonces ¿qué esperanzas depositaremos en el diseño de un experimento aleatorio que ponga de manifiesto la efectividad de este tipo de centros? La necesidad es la madre de la inventiva, y filósofos celebrados como Platón o Frank Zappa nos lo recuerdan. La naturaleza discreta de las políticas

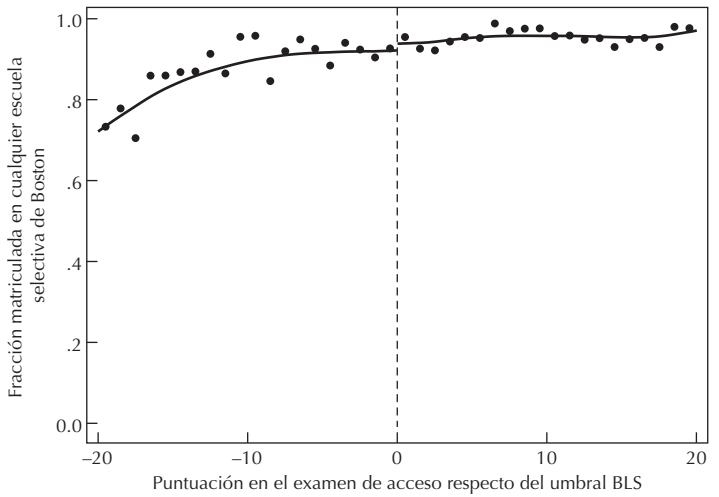



**Figura 4.7.** Matriculación en cualquier escuela selectiva de Boston. *Notas:* Esta figura representa las tasas de matriculación en cualquier escuela selectiva de Boston, en función de las notas en el examen de admisión, para quienes solicitaron plaza en la Boston Latin School (BLS) y obtuvieron calificaciones en el entorno del umbral. Las líneas continuas muestran valores estimados mediante regresión lineal local, con estimaciones separadas a cada lado del umbral (el cual se señala con la línea vertical discontinua).

de admisión de estos colegios crea un experimento natural. Entre los solicitantes con notas cercanas al umbral de admisión, el hecho de caer a la derecha o a la izquierda del corte podría comportarse de un modo tan bueno como si se hiciera al azar. Pero en este caso el experimento es sutil: aquí no se trata de que el umbral active un simple interruptor de encendido o apagado, sino que cambia de manera discontinua la naturaleza de la experiencia escolar, dado que algunos estudiantes admitidos eligen cursar estudios en otros lugares, mientras que muchos de los rechazados en una escuela selectiva acaban cursando estudios en otro centro de este mismo tipo. Cuando las discontinuidades alteran las probabilidades de tratamiento u otras características medias (la intensidad de tratamiento, por brevedad), en lugar de actuar sobre un simple interruptor de encendido o apagado, entonces resultan los diseños llamados de *RD difusa*.

### *RD difusa*

¿Cuál es exactamente el tratamiento en el caso de las escuelas selectivas? Las figuras 4.6-4.8, que se centran en solicitantes del centro BLS, nos ayudan a perfilar la respuesta. Los solicitantes del BLS, como todos los que aspiran a una plaza en las escuelas selectivas de Boston, se someten al Examen de Acceso de las Escuelas Independientes (ISEE, Independent School Entrance Exam). La muestra usada para construir estas figuras consiste en solicitantes con notas ISEE cercanas al umbral de acceso. Los puntos de las gráficas son promedios de la variable del eje *Y*, calculados para solicitantes cuyas notas ISEE entran en intervalos de un punto de anchura, mientras que la línea que pasa por entre los puntos presenta un ajuste obtenido al suavizar esos datos de la manera que se explica en la nota al pie.[6] La

[6] En estas figuras, la variable que determina la admisión es un promedio ponderado de la nota ISEE de cada solicitante y su GPA (Grade Point Average, la media de sus notas finales en cada curso), aunque nos refiramos a esta variable móvil como la «nota ISEE» para abreviar. Los puntos proceden de un método de suavizado conocido como regresión lineal local, que ajusta regresiones a muestras pequeñas definidas por un ancho de banda en el entorno de cada punto. Los valores suavizados son los valores estimados que se generan de este modo. Véanse los detalles en el estudio sobre el cual basamos nuestro tratamiento del tema: Atila Abdulkadiroglu, Joshua D. Angrist y Parag Pathak, «The Elite Illusion: Achievement Effects at Boston and New York Exam Schools», *Econometrica*, vol. 81, número 1, enero de 2014, páginas 137-196.

figura 4.6 muestra que la mayoría de los solicitantes que superan el examen se matricula en BLS, pero no todos.

BLS es la escuela selectiva más prestigiosa de Boston. ¿Adónde van los solicitantes que quedan por debajo del corte de BLS? La mayoría acude a la Boston Latin Academy, la institución venerable que sigue a BLS dentro de la jerarquía de centros selectivos de Boston. Este desplazamiento de matrículas se refleja en la figura 4.7, que representa las tasas de matriculación en cualquier centro selectivo de Boston en el entorno del umbral BLS. La figura 4.7 revela que la mayoría de estudiantes que no superan el umbral BLS acaba igualmente en otra escuela selectiva, de manera que las probabilidades de asistir a *alguna* escuela selectiva apenas se ven alteradas en torno al umbral BLS. Puede parecer, por tanto, que nos tenemos que conformar con un estudio de alcance limitado que compare el centro más selectivo, BLS, con la Boston Latin Academy, ligeramente menos selectiva, en lugar

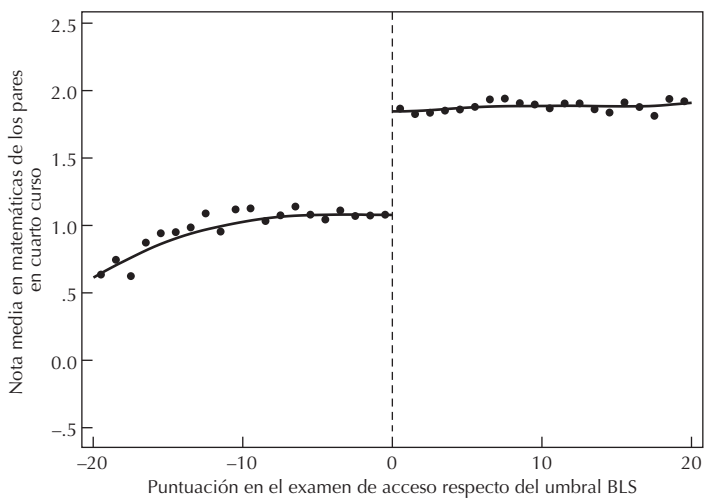



**Figura 4.8.** Calidad de los pares en el entorno del umbral BLS.
*Notas:* Esta figura representa la calidad media de los pares en séptimo curso, en función de las notas en el examen de admisión, para quienes solicitaron plaza en la Boston Latin School (BLS) y obtuvieron calificaciones en el entorno del umbral. La calidad de los pares se mide a través de las notas en matemáticas de cuarto curso de los compañeros de clase de séptimo. Las líneas continuas muestran valores estimados mediante regresión lineal local, con estimaciones separadas a cada lado del umbral (el cual se señala con la línea vertical discontinua).

de proceder a una evaluación más interesante de todo el esquema de escuelas selectivas.

¿O quizá no sea así? Una de las cuestiones más controvertidas en investigación educativa es la naturaleza de los efectos entre pares; es decir, si las capacidades de tus compañeros de clase ejercen algún efecto causal sobre tu propio aprendizaje. Si tienes la suerte de asistir a un centro de secundaria junto con otros estudiantes buenos, esto puede contribuir a tu éxito. Por otro lado, si te ves relegado a una escuela donde la mayoría del alumnado es mediocre, esto podría suponerte un freno. El efecto entre pares es importante para las políticas relacionadas con la asignación de centros, es decir, las normas y regulaciones que determinan a qué escuelas acuden los chicos. Por ejemplo, en muchas ciudades de Estados Unidos los alumnos acuden a los centros más cercanos a sus casas. Pero los estudiantes pobres, no blancos y con expedientes académicos peores suelen vivir alejados de la gente bien, apartados de los estudiantes que sacan buenas notas en barrios mayoritariamente blancos, de manera que la asignación de centros según el lugar de residencia podría reducir las posibilidades de lograr la excelencia para los chicos con pocos recursos o pertenecientes a minorías. Por eso, muchos distritos escolares trasladan a chicos lejos de donde viven, en un esfuerzo por incrementar la mezcla de estudiantes de distintos contextos y razas.

Los centros selectivos producen un experimento espectacular acerca de la influencia de la calidad de los pares. En concreto, los solicitantes que logran la admisión en uno de los centros selectivos de Boston cursan estudios en una escuela rodeados de pares de capacidades escolares mucho más elevadas que los solicitantes que quedan por debajo del umbral, incluso en los casos en que la alternativa es otro centro selectivo. La figura 4.8 documenta esta situación para los solicitantes de BLS. En este caso, la capacidad de los pares se mide mediante la nota en matemáticas de los compañeros de clase de los solicitantes en un examen efectuado en cuarto curso (dos años antes de que solicitaran el ingreso en los centros selectivos). Al igual que en la investigación sobre las escuelas *charter* tratada en el capítulo 3, las notas de examen que constan en esta figura se miden en unidades de desviación típica, donde una desviación típica se escribe en griego como $1\sigma$. Los solicitantes que logran el acceso a BLS estudian junto a compañeros de clase que sacan notas mucho más altas y disfrutan

de un salto en las calificaciones de sus pares en matemáticas de $0{,}8\sigma$, equivalente a la diferencia promedio en cuanto a capacidad de los pares entre la zona central de Boston y los ricos barrios del extrarradio. En el corazón de cualquier diseño de RD difusa se encuentra un cambio tan espectacular en la intensidad del tratamiento como el que tenemos aquí. La diferencia entre los diseños bruscos y los difusos consiste en que en el caso difuso los solicitantes que rebasan un umbral se exponen a un tratamiento más intenso, mientras que en el caso brusco el paso del umbral significa que el tratamiento pasa de estar del todo inactivo a activarse.

### *La técnica de RD difusa equivale a VI*

En el ritual de paso de la regresión, los científicos sociales de todo el mundo vinculan los logros de los estudiantes con la capacidad promedio de sus compañeros de clase. Estas regresiones revelan de manera fiable una asociación fuerte entre los éxitos académicos de los estudiantes y la calidad de sus pares. Si se aplica a los solicitantes de acceso en todas las escuelas selectivas de Boston, una regresión de las notas de matemáticas en séptimo curso sobre las notas medias de cuarto curso de sus compañeros de clase de séptimo genera un coeficiente con valor cercano a un cuarto. Este posible efecto de los pares surge del modelo de regresión

$$Y_i = \alpha_0 + \rho D_i + \beta_0 R_i + e_{0i}. \tag{4.6}$$

donde $Y_i$ es la nota en matemáticas del estudiante $i$ en séptimo curso, $X_i$ la nota en matemáticas del estudiante $i$ en cuarto curso, y $\overline{X}_{(i)}$ representa la nota media en matemáticas de todos los compañeros de clase del estudiante $i$ en cuarto curso (el subíndice *(i)* nos recuerda que el estudiante $i$ no se incluye a la hora de calcular el resultado medio de sus pares). La estimación que resulta para el coeficiente de calidad de los pares ($\theta_1$) está alrededor de 0,25, lo que significa que si se incrementara en una desviación típica la habilidad media de los pares en la escuela secundaria, medida a través de las notas que sacaron en primaria y controlando las notas de primaria del propio estudiante, resultaría un incremento de $0{,}25\sigma$ en los resultados en la escuela secundaria.

Los progenitores y el profesorado tienen una fuerte intuición de que «los pares importan», así que la fuerte asociación positiva entre los resultados de los estudiantes y los de sus compañeros de clase suena a verdadera. Pero es poco probable que esta regresión ingenua entre pares se pueda interpretar en términos causales, por la sencilla razón de que los estudiantes que se forman juntos tienden a parecerse entre sí por muchos motivos. Los cuatro hijos de quien esto escribe, por ejemplo, fueron estudiantes de éxito desde una edad temprana, como sus padres, y tuvieron la fortuna de asistir a escuelas junto a muchos muchachos de familias parecidas. Como el contexto familiar no se mantiene fijo en regresiones como la ecuación (4.6), la asociación observada entre los estudiantes y sus compañeros de clase refleja, sin lugar a dudas, algunas de estas influencias compartidas. Para salir del atolladero causal resultante, nos gustaría asignar los estudiantes al azar entre todo un abanico de distintos grupos de pares.

¡Los centros selectivos acuden al rescate! La figura 4.8 muestra la sensible diferencia en cuanto a capacidad de los pares inducida por la admisión en BLS, con un salto de $\frac{4}{5}$ de desviación típica en la capacidad en el umbral de admisión. El salto en calidad de los pares en los umbrales de admisión de las escuelas selectivas se produce (por diseño) a partir de la mezcla de estudiantes que se matriculan en este tipo de centros. Esto es justo lo que un econometrista encargaría si pidiera un experimento ideal sobre este asunto (esta mejora en la calidad de los pares es también la que hace que muchas familias anhelen el sueño de que sus hijos consigan plaza en una escuela selectiva). Además, en tanto que la calidad de los pares se incrementa en el umbral, no se aprecian saltos similares al comparar a un lado y otro las demás variables relacionadas con las capacidades propias de los solicitantes, como la motivación y el contexto familiar (las fuentes de sesgos de selección por las que nos solemos preocupar). Por ejemplo, no hay salto en las calificaciones de los propios solicitantes en la escuela primaria. Los pares cambian de manera discontinua en el umbral de admisión, pero no lo hacen las características propias de los solicitantes de acceso en las escuelas selectivas.[7]

[7] Esto se documenta en Abdulkadiroglu *et al.*, «The Elite Illusion», *Econometrica*, 2014.

Pero pese a los anhelos y sueños, y a los resultados de nuestra regresión ingenua acerca del efecto de los pares (ecuación [4.6]), el experimento de los centros selectivos arroja dudas sobre la idea de que exista el efecto causal debido a los pares sobre los resultados académicos de los solicitantes de las escuelas selectivas de Boston. La semilla de la duda la planta la figura 4.9, que representa los resultados en matemáticas de quinto y sexto cursos (exámenes realizados tras uno o dos años de estudios secundarios) frente a las notas ISEE (la variable móvil) para solicitantes con notas en el entorno del umbral BLS. Los solicitantes que logran la admisión se exponen a un grupo de pares mucho más capaces, pero esta exposición no induce un salto equivalente en los resultados de estos solicitantes en los estudios secundarios.

Como en la ecuación (4.2), el tamaño del salto en la figura 4.9 se puede estimar si se estima una ecuación como

$$Y_i = \alpha D_i + \beta_0 R_i + e_{0i}. \tag{4.7}$$

Aquí $D_i$ es una variable binaria que marca a los solicitantes que logran el acceso, mientras que $R_i$ es la variable móvil que determina si se entra o no. En una muestra de solicitantes de acceso en BLS de séptimo curso, donde $Y_i$ representa las notas de matemáticas en secundaria como en las figuras, esta regresión genera una estimación de –0,2, con un error típico de 0,20, un cero estadístico de libro.

¿Cómo interpretar esta estimación de $\rho$? Por supuesto, ¡mediante la lupa de la primera etapa correspondiente! La ecuación (4.7) es la forma reducida de un esquema de MC2E donde la variable endógena sería la calidad media de los pares, $\bar{X}_{(i)}$. La ecuación de la primera etapa vinculada a esta forma reducida es:

$$\bar{X}_{(i)} = \alpha_1 + \phi D_i + \beta_1 R_i + e_{0i}, \tag{4.8}$$

donde el parámetro $\phi$ evalúa el salto en calidad media de los pares inducido por la oferta de plaza en la escuela selectiva. Este es el salto que se muestra en la figura 4.8, un valor estimado con precisión e igual a $0{,}80\sigma$.

La última pieza del esquema de MC2E es la relación causal de interés, la segunda etapa MC2E. En este caso, la segunda etapa capta

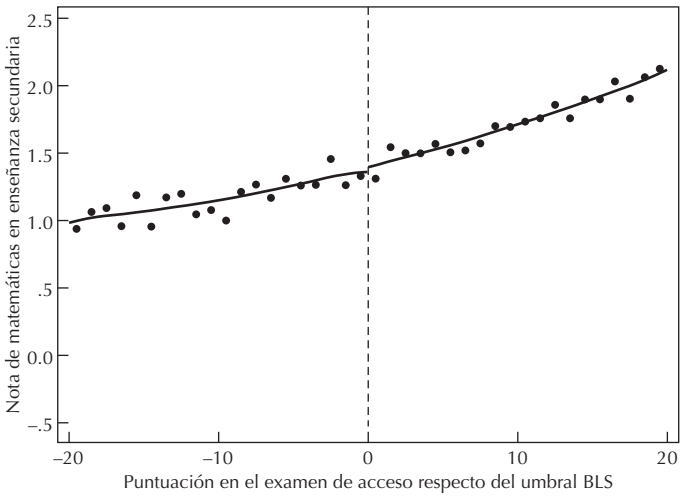



**Figura 4.9.** Notas en matemáticas en el entorno del umbral BLS. *Notas:* Esta figura representa las notas de matemáticas en los cursos séptimo y octavo de los solicitantes de plaza en la Boston Latin School (BLS), en función de las notas en el examen de admisión, para quienes obtuvieron calificaciones en el entorno del umbral. Las líneas continuas muestran valores estimados mediante regresión lineal local, con estimaciones separadas a cada lado del umbral (el cual se señala con la línea vertical discontinua).

el efecto de la calidad de los pares sobre las notas de matemáticas en los cursos séptimo y octavo. Como siempre, la segunda etapa incluye las mismas variables de control que aparecen en la primera. Esto conduce a una ecuación de segunda etapa que se puede escribir como

$$Y_i = \alpha_2 + \lambda \hat{X}_{(i)} + \beta_2 R_i + e_{2i}, \tag{4.9}$$

donde $\lambda$ es el efecto causal de la calidad de los pares, y la variable $\hat{X}_{(i)}$ representa el valor estimado de la primera etapa, resultado de estimar la ecuación (4.8).

Obsérvese que la ecuación (4.9) hereda una covariable de la primera etapa y de la forma reducida, la variable móvil, $R_i$. Por otro lado, la variable binaria del salto, $D_i$, queda excluida de la segunda etapa porque es el instrumento que hace funcionar la maquinaria de los MC2E. Tengamos en cuenta que hemos admitido que en el entorno del umbral de admisión, y si se estiman los efectos de la variable móvil

por medio de un control lineal, las notas de acceso para la escuela selectiva no ejercen ningún efecto directo sobre las notas en los exámenes, sino que repercute en los resultados académicos, si es que lo hace, tan sólo a través de la calidad de los pares. Esta suposición equivale a la importante restricción de exclusión del sistema VI en este contexto.

La estimación MC2E de $\lambda$ en la ecuación (4.9) vale −0,023, con un error típico de 0,132.[8] Como la estimación de la forma reducida es cercana a cero y no difiere de esa cifra de manera significativa, también es nula la correspondiente estimación MC2E. Esta estimación también dista mucho de la estimación de $0{,}35\sigma$ generada por la estimación por MCO de la regresión ingenua de los efectos de los pares, ecuación (4.6). Por otra parte, ¿quién puede afirmar que lo único importante de las escuelas selectivas es la calidad de los pares? La restricción de exclusión requiere atenerse a un canal causal concreto. Pero este canal no tiene por qué ser el único relevante en la práctica.

Un rasgo distintivo del contexto de las escuelas selectivas, junto a la calidad de los pares, es la composición racial. En el ambiente de los colegios públicos de Boston, que acogen sobre todo a minorías, los centros selectivos brindan la oportunidad de acudir a un colegio con una población más diversa, entendiendo ahora la diversidad como una proporción mayor de compañeros de clase blancos. La disolución de los sistemas educativos estadounidenses con segregación racial, impuesta por ley, vino motivada por un esfuerzo para mejorar los resultados académicos. En 1954, el Tribunal Supremo de Estados Unidos emitió aquella famosa declaración: «Las instalaciones educativas separadas son intrínsecamente contrarias a la igualdad», con lo que sentó las bases para que los juzgados condujeran a un incremento del equilibrio racial en los colegios públicos. ¿Incrementar el equilibrio racial mejora los resultados académicos? Los centros selectivos son relevantes en el debate de la integración racial porque la admisión en estas escuelas aumenta marcadamente la exposición a compañeros

[8] Este error típico está agrupado a nivel de solicitante. Como se explica en el apéndice del capítulo 5, usamos errores típicos robustos por grupos para tener en cuenta el hecho de que los datos contienen observaciones correlacionadas (en este caso, las notas de séptimo y octavo cursos para cada solicitante BLS están correlacionadas).

de clase blancos. A la vez, sabemos que si la calidad de los pares, $\overline{X}_{(i)}$, se sustituye por una medida de la proporción de estudiantes blancos, esto también genera un coeficiente de segunda etapa nulo, como consecuencia del hecho de que la forma reducida subyacente no se ve alterada por la elección del canal causal.

Las universidades selectivas pueden ser distintas también en otros sentidos, porque quizá atraigan a profesores mejores, u ofrezcan mejores cursos de orientación universitaria (a nivel de universidad) que los centros públicos no selectivos. Conviene subrayar que hay otros aspectos del entorno escolar que pueden cambiar por el corte en el umbral de admisión de las escuelas selectivas, como los recursos con que cuenta el centro, y que pueden resultar beneficiosos. Esto a su vez sugiere un efecto positivo para cualesquiera variables omitidas asociadas a las estimaciones MC2E de los efectos de la calidad de los pares. Dado que las variables omitidas con efectos positivos probablemente estén correlacionadas positivamente con la oferta de plaza en el centro selectivo, la estimación MC2E que usa las calificaciones en los centros selectivos como instrumento para la calidad de los pares será, en todo caso, demasiado grande en comparación con el efecto aislado de los pares que estamos estudiando. Lo cual torna aún más sorprendente que este efecto estimado resulte ser nulo.

Como en cualquier estudio IV, la RD difusa requiere valoraciones sólidas de los canales causales a través de los cuales los instrumentos inciden en los resultados. En la práctica hay muchos canales que pueden transmitir los efectos causales, y en esos casos exploramos las alternativas. Es verosímil que los canales que midamos no tengan por qué ser los únicos relevantes. El viaje causal nunca termina, continuamente surgen preguntas nuevas. Pero el esquema difuso que usa RD para generar instrumentos no es menos útil a este respecto.

Maestro Stevefu: Resume para mí la RD, Pequeño Saltamontes.

Pequeño Saltamontes: Los diseños de RD explotan los cambios bruscos en el estado de tratamiento que se producen cuando el tratamiento se fija mediante un umbral.

Maestro Stevefu: ¿Y es la RD tan buena como un ensayo aleatorio?

Pequeño Saltamontes: La RD exige conocer la relación entre la variable móvil y los resultados potenciales en ausencia de tratamiento. Tenemos que controlar esta relación cuando se usan las discontinuidades para identificar efectos causales. Los ensayos aleatorios no requieren tales controles.

Maestro Stevefu: ¿Cómo saber si la estrategia de control es adecuada?

Pequeño Saltamontes: Nunca se tiene esa certeza, maestro. Pero se refuerza la confianza en las conclusiones causales cuando las estimaciones RD permanecen similares aunque cambien los detalles del modelo RD.

Maestro Stevefu: ¿Y qué decir de lo brusco frente a lo difuso?

Pequeño Saltamontes: La versión brusca se da cuando el tratamiento mismo se activa o desactiva al pasar el umbral. La versión difusa corresponde a la situación en la que el paso del umbral induce un salto en la probabilidad o la intensidad del tratamiento. En los diseños difusos una variable que marca si se ha rebasado el umbral se convierte en un instrumento; el diseño difuso se analiza por medio de MC2E.

Maestro Stevefu: Te aproximas al umbral de la maestría, Pequeño Saltamontes.

## Maestros de la econometría: Donald Campbell

El relato de la RD lo narraron por vez primera los psicólogos Donald L. Thistlethwaite y Donald T. Campbell, quienes emplearon RD en 1960 para evaluar el impacto de los premios Becas Nacionales al Mérito sobre las carreras y actitudes de los premiados.[9] Como muchos de nuestros lectores quizá sepan, el programa de Becas Nacionales al Mérito es un proceso en varias rondas, al final de las cuales unos cuantos miles de graduados de secundaria reciben una beca universitaria. La selección se basa en los resultados de los solicitantes en los exámenes PSAT y SAT, que son los exámenes de acceso a la universi-

[9] Donald L. Thistlethwaite y Donald T. Campbell, «Regression-Discontinuity Analysis: An Alternative to the Ex Post Facto Experiment», *Journal of Educational Psychology*, vol. 51, número 6, diciembre de 1960, páginas 309-317.

dad por los que pasa la mayoría de estudiantes que solicita ingresar en la universidad.

Los candidatos que triunfan en la competición de las Becas Nacionales al Mérito consiguen unas notas PSAT por encima de un umbral (y validan estas notas PSAT si aprueban el SAT, que se hace después). Entre estos, algunos reciben una beca concedida por el comité de filtrado de la competición, mientras que a los demás se les otorga un certificado de mérito. Los estudiantes que reciben este certificado, los llamados finalistas del Mérito Nacional, tienen motivos para alegrarse, porque sus nombres se distribuyen entre las facultades, universidades y otros patrocinadores de becas. Las facultades que cuentan con muchos finalistas del Mérito Nacional entre sus estudiantes también suelen anunciar esta circunstancia. Thistlethwaite y Campbell se preguntaron si un reconocimiento como el de ser finalista del Mérito Nacional ejerce algunas consecuencias duraderas entre quienes lo reciben.

En un trabajo anterior basado en métodos de emparejamiento (como los tratados en el capítulo 2), Thistlethwaite estimó que los solicitantes que lograron un certificado de mérito tenían 4 puntos porcentuales más de probabilidad de aspirar a convertirse en profesores universitarios o investigadores que si no lo hubieran logrado.[10] Pero un diseño RD que explotaba las discontinuidades en el entorno del umbral PSAT para obtener el certificado de mérito mostró una estimación estadísticamente no significativa de este resultado, de tan sólo dos puntos. Reproducimos como figura 4.10 la gráfica que acompañaba a este hallazgo. El reconocimiento público por sí mismo parece ejercer un efecto escaso en la elección de carrera o en los planes sobre qué estudiar en la universidad.

Donald Campbell es recordado no sólo por inventar la RD, sino también por su ensayo de 1963, «Experimental and Quasi-Experimental Designs for Research on Teaching» [«Diseños experimentales y casi experimentales para la investigación sobre la enseñanza»], escrito con Julian C. Stanley y publicado más tarde en forma de libro. El ensayo de Campbell y Stanley constituyó una exploración pionera de los métodos econométricos explicados en el presente capítulo y en

[10] Donald L. Thistlethwaite, «Effects of Social Recognition upon the Educational Motivation of Talented Youths», *Journal of Educational Psychology*, vol. 50, número 3, 1959, páginas 111-116.

el siguiente de este libro. Una actualización posterior escrita con Thomas D. Cook sigue siendo una referencia importante hasta hoy día.[11]

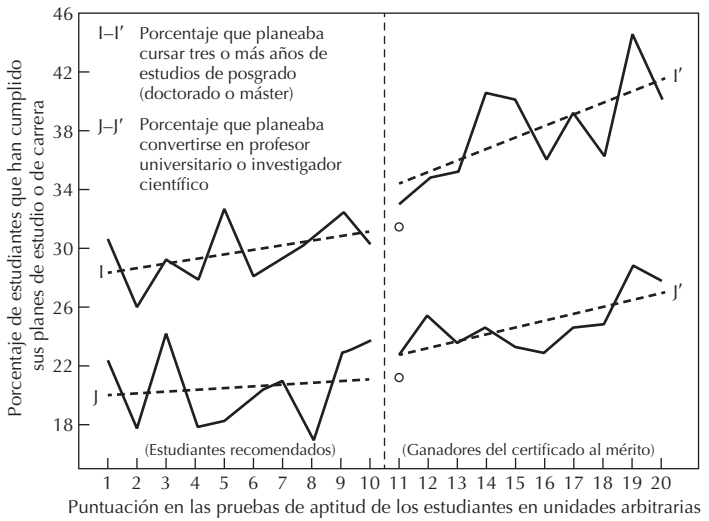



**Figura 4.10.** RD visual de Thistlethwaite y Campbell.
*Notas:* Esta figura representa, para quienes realizaron los exámenes PSAT, los planes sobre estudios de posgrado (línea I-I') y una medida de los planes de carrera (línea J-J') frente a la variable móvil que determina el reconocimiento de Mérito Nacional.

[11] Donald T. Campbell y Julian C. Stanley, «Experimental and Quasi-Experimental Designs for Research on Teaching», capítulo 5 de Nathaniel L. Gage (ed.), *Handbook of Research on Teaching*, Rand McNally, 1963; y Donald T. Campbell y Thomas D. Cook, *Quasi-Experimentation: Design and Analysis Issues for Field Settings*, Houghton Mifflin, 1979.

5

# Diferencias en diferencias

Maestro Kan: Si al construir una casa un carpintero golpea un clavo y ve que no sirve porque está doblado, ¿pierde el carpintero la fe en todos los clavos y deja de construir? Lo mismo sucede con el trabajo empírico.

*Kung Fu*, temporada 1, episodio 7

## *Nuestro camino*

A veces cuesta encontrar variables instrumentales creíbles o discontinuidades espectaculares en alguna normativa, así que hay que tener otras herramientas disponibles en el taller econométrico. El método de las *diferencias en diferencias* (DD) reconoce que, en ausencia de una asignación aleatoria, es muy probable que los grupos de control y de tratamiento difieran por muchos motivos. Sin embargo, a veces la población ulteriormente sujeta a tratamiento y las poblaciones de control siguen caminos paralelos en ausencia de tratamiento. Cuando esto sucede, una evolución divergente una vez aplicado el tratamiento, respecto de la tendencia previa puede reflejar el impacto del tratamiento. Ilustraremos el método DD mediante un estudio de los efectos de la política monetaria sobre la quiebra de algunos bancos durante la Gran Depresión. También volveremos a visitar la edad mínima legal para el consumo de alcohol en Estados Unidos (MLDA).

## 5.1 El experimento Misisipi

En los albores de la mayor catástrofe económica de la historia de Estados Unidos (la Gran Depresión) imperaba un gran optimismo en los salones de las grandes finanzas. El eslogan de Caldwell and Company «Confiamos en el Sur»,[1] reflejaba la confianza de un imperio financiero regional. Caldwell dirigía la mayor cadena de bancos del sur de Estados Unidos, con sede en Nashville, en la década de 1920, y poseía también muchos otros negocios de carácter no bancario. Rogers Caldwell, conocido como «el J.P. Morgan del Sur», vivía a lo grande en una finca que albergaba su propio establo de caballos campeones de pura raza. Pero he aquí que en noviembre de 1930 una gestión deficiente y el derrumbe del mercado de valores de octubre de 1929 tumbaron el imperio de Caldwell. En cuestión de días, el colapso de Caldwell arrastró consigo sus redes bancarias, estrechamente entretejidas, en Tennessee, Arkansas, Illinois y Carolina del Norte. La crisis de Caldwell fue el heraldo de toda una riada de quiebras de bancos por todo el país.

El negocio bancario se basa en la fiabilidad y la confianza. Los bancos prestan a las empresas y a las personas, con la expectativa de que la mayoría de los créditos se paguen cuando llegue la fecha acordada. Quienes contratan depósitos confían en poder retirar sus fondos cuando lo deseen. A pesar de ello, los bancos tienen menos dinero en metálico del que necesitarían para abonar todos sus depósitos, porque la mayor parte de esos depósitos los han prestado. El consecuente desfase de vencimientos no supone ningún problema en circunstancias normales, cuando son pocos los clientes que solicitan reembolsos en un día cualquiera.

Pero cuando se produce una quiebra de la confianza, el sistema bancario se hunde. En la década de 1930, si tu banco cerraba, lo más probable era que tus ahorros desaparecieran con él. Aunque el balance de hipotecas y préstamos de tu banco parezca sólido, no querrás ser el último cliente en pedir el reembolso de tu dinero. Y cuando veas a otros clientes en pánico y retirando sus fondos, lo mejor que puedes hacer es entrar en pánico tú también. Así empieza la quiebra de un banco.

El colapso de Caldwell influyó en la confianza de los clientes de los bancos por todo el sur de Estados Unidos y precipitó una avalancha de

[1] «We Bank on the South». *(N. de la T.)*

retiradas de depósitos de los bancos del estado de Misisipi en diciembre de 1930. Al principio los depósitos bancarios de Misisipi fueron disminuyendo lentamente, pero el 19 de diciembre se abrieron las compuertas cuando los ahorradores sucumbieron al pánico. Ese día el Departamento de Bancos del estado de Misisipi cerró tres bancos. Dos más dejaron de operar al día siguiente, y otros 29 los siguieron en los seis meses posteriores. Este pánico regional de 1930 fue sólo uno de los muchos otros que estaban por venir. Las quiebras bancarias de la era de la Depresión alcanzaron su máximo en 1933, con más de 4.000 quiebras por todo el país.

La ciencia económica ha intentado durante mucho tiempo comprender si la política monetaria contribuyó a la Gran Depresión, de qué modo pudo hacerlo, y si una intervención monetaria más agresiva podría haber contenido el desmoronamiento financiero y la caída libre de la economía que sucedió en aquellos días oscuros. Las lecciones de la era de la Depresión pueden ayudarnos a entender el presente. Aunque los mercados financieros de ahora sean más complejos, los pilares de las finanzas siguen siendo iguales que entonces: los bancos toman prestado a corto y prestan, normalmente con vencimientos a más largo plazo, y confían en ser capaces de disponer del dinero en metálico (en la jerga bancaria, la «liquidez») que necesiten para cumplir con sus obligaciones a medida que venzan.

Tenemos la mala suerte de vivir unos tiempos económicos interesantes. El año 2008 vio agitarse el sistema financiero estadounidense por un colapso del mercado de garantías hipotecarias, seguido por una crisis de la deuda soberana europea que dio comienzo a finales de 2009. Carmen Reinhart y Kenneth Rogoff han descrito hace poco las crisis financieras acaecidas desde el siglo xiv, y defienden que todas ellas comparten la misma anatomía. La semejanza aparente de todos esos episodios nos lleva a preguntarnos si podrían evitarse o, cuando menos, si cabría mitigar sus efectos. Milton Friedman y Anna Schwartz, en su magistral historia monetaria de Estados Unidos, de 1960, convencieron a muchos economistas de que una comprensión certera de los efectos de la política monetaria constituye la clave para responder esta cuestión.[2]

[2] Carmen Reinhart y Kenneth Rogoff, *This Time Is Different: Eight Centuries of Financial Folly*, Princeton University Press, 2009; y Milton Friedman y Anna Schwartz, *A Monetary History of the United States, 1867-1960*, Princeton University Press, 1963.

### *Un Misisipi, dos Misisipis*

Cuando surge un problema bancario, quienes definen las políticas pueden, o bien abrir el flujo del crédito, o bien cerrar el grifo. Friedman y Schwartz sostenían que la Reserva Federal (el banco central de Estados Unidos) adoptó un actitud demencial al restringir el crédito a medida que se desarrollaba la Gran Depresión. El dinero fácil podría haber permitido a los bancos satisfacer las demandas crecientes de retirada urgente de depósitos y apaciguar así el pánico de los clientes. Prestando con liberalidad a los bancos con problemas, el banco central tiene el poder de calmar una crisis de liquidez y de evitar un rescate.

Pero ¿quién puede decir cuándo una crisis es tan sólo una crisis de confianza? Algunas crisis son reales. Los balances bancarios pueden estar tan enfermos debido a deudas fallidas que no haya ningún aporte temporal de liquidez capaz de curarlos. Al fin y al cabo, los bancos no pierden la liquidez de froma aleatoria. Al contrario, los directivos de los bancos efectúan préstamos que, o bien fracasan, o son fructíferos. Inyectar fondos del banco central en bancos malos puede significar tirar dinero bueno encima del malo. En esos casos puede ser mejor declarar la bancarrota y confiar en que se haga una distribución ordenada de los activos que queden.

Dar apoyo a los bancos malos resucita el fantasma de lo que en economía se denomina *riesgo moral*. Si los banqueros saben que el banco central prestará dinero barato cuando haya falta de liquidez, tienen menos razones para adoptar precauciones con el fin de evitar una posible crisis. En 1873 el redactor jefe de *The Economist*, Walter Bagehot, describió este riesgo con estas palabras:

> Si los bancos son malos, sin duda lo seguirán siendo y se volverán peores si el gobierno los sostiene y los favorece. La máxima cardinal es que cualquier ayuda actual a un banco malo constituye la manera más certera de impedir el establecimiento de un banco bueno en el futuro.[3]

Bagehot era un darwinista social confeso que creía que los principios evolucionistas sirven para los asuntos sociales del mismo modo que para la biología. ¿Qué decisión tiene más probabilidades de favorecer

[3] Del capítulo IV.4 de Walter Bagehot, *Lombard Street: A description of the Money Market*, Henry S. King and Co., 1873.

un final feliz ante una dificultad económica, un aporte de liquidez o dejar que sobrevivan los bancos mejor adaptados? Como siempre, a los maestros de la econometría les gustaría esclarecer esta cuestión con un experimento aleatorio. Hemos cursado una petición de financiación para un experimento de este tipo, y la están evaluando; podemos asegurar que comunicaremos los resultados si tal petición tuviera éxito. Pero entretanto no nos queda más remedio que conocer los efectos de la política monetaria a partir de la historia de las crisis bancarias y de las respuestas que se dieron a las mismas.

Por suerte para este programa de investigación, el sistema de la Reserva Federal de Estados Unidos está organizado en 12 distritos, cada uno de ellos regido por un Banco de Reserva Federal regional. Los gobernadores de cada distrito gozaban de bastante independencia en la era de la Depresión. El sexto distrito, dirigido desde Atlanta, favorecía prestar dinero a los bancos con problemas. En cambio, el octavo distrito, dirigido desde San Luis, seguía una filosofía conocida como la doctrina «Real Bills», que sostenía que había que restringir el crédito en el curso de una recesión. Por fortuna para la investigación en materia de política monetaria, la frontera entre los distritos sexto y octavo atraviesa de este a oeste el centro del estado de Misisipi (las fronteras entre distritos se decidieron de acuerdo a criterios de población en 1913, cuando nació el sistema de la Reserva Federal). Esta frontera define un experimento natural dentro del estado, del cual podemos beneficiarnos.

Los maestros Gary Richardson y William Troost analizaron la situación monetaria en un Misisipi dividido en dos.[4] Como cabría esperar de sus dos formas de entender la política monetaria, las Reservas Federales de Atlanta y de San Luis reaccionaron de maneras muy distintas a la crisis de Caldwell. En las dos semanas siguientes al colapso de Caldwell, Atlanta había incrementado los préstamos a los bancos en un 40% en el sexto distrito. En ese mismo periodo los préstamos a los bancos del octavo distrito, el de San Luis, cayeron en casi un 10%.

[4] Gary Richardson y William Troost, «Monetary Intervention Mitigated Banking Panics during the Great Depresssion: Quasi-Experimental Evidence from a Federal District Border, 1929-1933», *Journal of Political Economy*, vol. 117, número 6, diciembre de 2009, páginas 1031-1073. Las cifras de este apartado proceden de los datos de Richardson y Troost.

El experimento de Richardson y Troost toma el distrito octavo como grupo de control, allí donde la política consistió en no hacer nada o incluso en restringir los préstamos, mientras que el sexto distrito se considera el grupo de tratamiento, donde se siguió la política de aumentar el crédito. El primer resultado es el número de bancos que seguían operando en cada distrito el 1 de julio de 1931, unos ocho meses después del comienzo de la crisis. Ese día había 132 bancos abiertos en el distrito octavo y 121 en el sexto, una diferencia de 11 entre uno y otro. Esto parecería indicar que facilitar liquidez había sido contraproducente. Pero mirémoslo con cuidado: los distritos sexto y octavo eran parecidos, pero no idénticos. Esto se aprecia en el hecho de que el número de bancos que operaban en los dos distritos difería notablemente ya el 1 de julio de 1930, mucho antes de la crisis de Caldwell, con 135 bancos abiertos en el distrito sexto y 165 en el octavo. Analizamos el experimento de Misisipi mediante una técnica llamada *diferencias en diferencias*, o DD para abreviar, para tener en cuenta esta disparidad entre distritos en el periodo pretratamiento.

### *Mundos paralelos*

Sea $Y_{dt}$ el número de bancos abiertos en el distrito $d$ en el año $t$, donde el subíndice $d$ nos dice si estamos considerando los datos del distrito sexto o del octavo, y $t$ distingue entre los datos de 1930 (antes de la crisis de Caldwell) y los de 1931 (después de esa crisis). La estimación DD ($\delta_{DD}$) del efecto de proporcionar liquidez en el distrito sexto es:

$$\begin{aligned}\delta_{DD} &= (Y_{6,\,1931} - Y_{6,\,1930}) - (Y_{8,\,1931} - Y_{8,\,1930}) \\ &= (121 - 135) - (132 - 165) \\ &= -14 - (-33) = 19. \end{aligned} \tag{5.1}$$

En lugar de comparar el número de bancos abiertos en los distritos sexto y octavo después de la crisis de Caldwell, la técnica DD contrasta la variación en el número de bancos que operaban en los dos distritos.

Al comparar los cambios en lugar de comparar los números existentes se tiene en cuenta el hecho de que en el periodo pretratamien-

to el distrito octavo tenía más bancos abiertos que el sexto. Para ver esto obsérvese que podemos obtener el mismo resultado básico DD de este modo:

$$\begin{aligned}\delta_{DD} &= (Y_{6,\,1931} - Y_{8,\,1931}) - (Y_{6,\,1930} - Y_{8,\,1930}) \\ &= (121 - 132) - (135 - 165) \\ &= -11 - (-30) = 19. \end{aligned} \tag{5.2}$$

Esta versión del cálculo DD toma la diferencia pretratamiento entre los distritos sexto y octavo, y la sustrae a la diferencia postratamiento, ajustando por el hecho de que los dos distritos eran diferentes de partida. Las estimaciones DD indican que prestar dinero a los bancos en apuros mantuvo abiertos muchos de ellos. En concreto, la decisión tomada en Atlanta parece haber salvado 19 bancos, más del 10% de los que operaban en 1930 en el sexto distrito de Misisipi.

La figura 5.1 ilustra la lógica DD, donde se representa el número de bancos en los distritos sexto y octavo en 1930 y 1931, uniendo con líneas continuas los datos de los dos periodos. La figura 5.1 resalta el hecho de que, aunque fracasaran bancos en los dos distritos, esto ocurrió de manera más acentuada en el octavo.

La herramienta DD equivale a comparar las pendientes o las tendencias entre distritos. La línea discontinua de la figura 5.1 representa el resultado contrafactual que yace en el centro del diseño del procedimiento DD: esta línea nos informa de lo que habría ocurrido en el distrito sexto en caso de que la situación hubiera evolucionado como lo hizo en el octavo. El hecho de que la línea continua del distrito sexto descienda de un modo mucho más suave que esta gráfica contrafactual constituye un indicio de la efectividad del aporte de liquidez. Las 19 quiebras bancarias descubiertas por nuestro cálculo DD son la diferencia entre lo que ocurrió de verdad y lo que habría ocurrido si la actividad bancaria en ambos distritos se hubiera desarrollado en paralelo.

La previsión contrafactual DD procede de un supuesto fuerte pero fácil de describir: *las tendencias comunes*. En el experimento de Misisipi la técnica DD presupone que, en caso de que no se aplicaran políticas diferentes, lo que habría cabido esperar en el distrito sexto es la tendencia del distrito octavo. Aunque no sea un supuesto inocuo, plantear unas tendencias idénticas parece un punto de partida razonable,

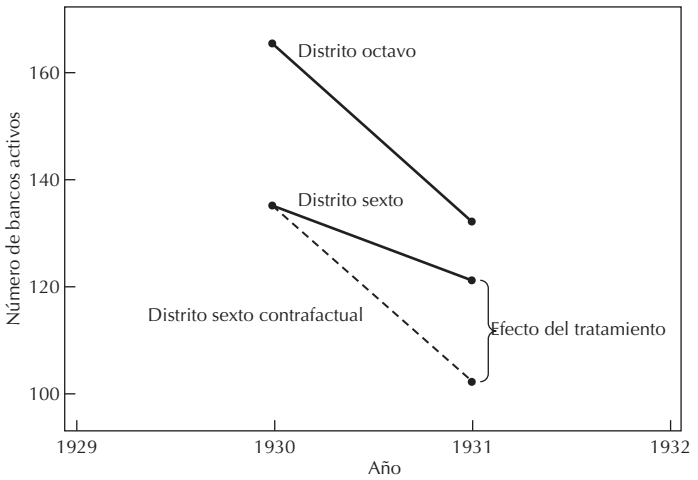



**Figura 5.1.** Quiebras bancarias en los distritos sexto y octavo de la Reserva Federal.

*Notas:* Esta figura muestra el número de bancos activos en los distritos sexto y octavo de la Reserva Federal en Misisipi en los años 1930 y 1931. La línea discontinua representa la evolución contrafactual del número de bancos en el distrito sexto si en este periodo de tiempo hubiera quebrado en ese distrito la misma cantidad de bancos que en el distrito octavo.

uno que tiene en cuenta las diferencias anteriores al tratamiento. Si se dispone de más datos, entonces se puede explorar, poner a prueba y, eventualmente, relajar este supuesto.

La figura 5.2 aporta indicios a favor del supuesto de una tendencia idéntica en los dos distritos de la Reserva Federal en Misisipi. Los indicios se presentan en la forma de una serie temporal más amplia sobre la actividad bancaria. Antes de 1931 la Gran Depresión aún no había golpeado con dureza a Misisipi. Las políticas regionales de la Reserva Federal en los dos distritos fueron, además, similares durante este periodo algo más tranquilo. Las quiebras bancarias se movieron casi en paralelo en los dos distritos entre 1929 y 1930, con un ligero descenso en el número de bancos en ambas zonas, lo que concuerda con la hipótesis de una misma tendencia en los periodos sin tratamiento. La figura 5.3 añade un sexto distrito contrafactual, el que se deduciría de extrapolar las tendencias del distrito octavo al sexto para los años posteriores a 1930. El contraste entre la actividad bancaria en el sexto distrito real y el contrafactual cambia poco hasta 1943.

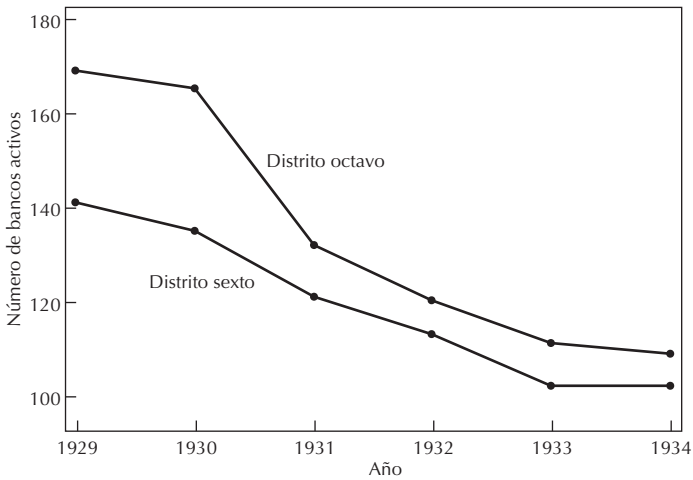



**Figura 5.2.** Tendencias de las quiebras bancarias en los distritos sexto y octavo de la Reserva Federal.
*Notas:* Esta figura muestra el número de bancos activos en los distritos sexto y octavo de la Reserva Federal en Misisipi entre los años 1929 y 1934.

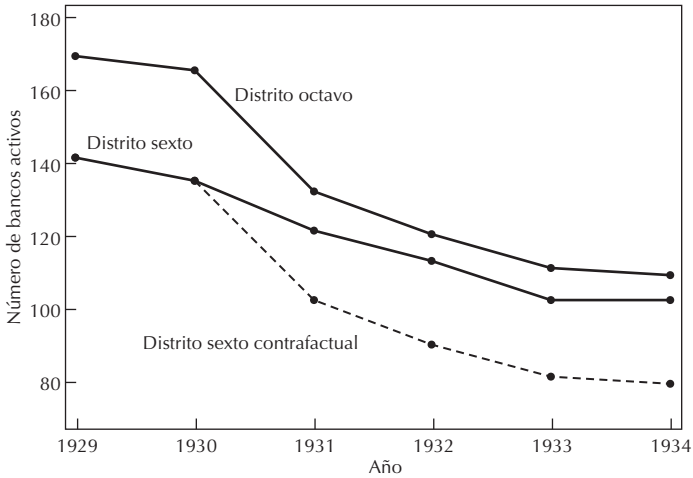



**Figura 5.3.** Tendencias de las quiebras bancarias en los distritos sexto y octavo de la Reserva Federal, y estimación DD contrafactual para el distrito sexto.
*Notas:* Esta figura añade los resultados DD contrafactuales a los datos bancarios de la figura 5.2. La línea discontinua traza la evolución contrafactual del número de bancos en el distrito sexto si el número de quiebras bancarias en ese distrito a partir de 1930 hubiera sido el mismo que en el distrito octavo.

Como en la figura 5.1, la caída de la actividad bancaria relativamente abrupta en el distrito octavo tras el colapso de Caldwell queda bien clara en las figuras 5.2 y 5.3. Pero estas gráficas documentan algo más. A principios de julio de 1931 San Luis abandonó la línea dura y empezó a prestar dinero abiertamente a los bancos en apuros. En otras palabras, desde 1931 la política de la Reserva Federal en los dos distritos volvió a ser similar, con ambas oficinas aportando liquidez a manos llenas. Además, mientras que la depresión distaba mucho de haber pasado en 1932, la crisis Caldwell había amainado y las retiradas de fondos habían recuperado los niveles anteriores a la crisis. Como ya había una disposición común de las dos oficinas regionales de la Reserva Federal a prestar el dinero que fuera necesario, las tendencias de la actividad bancaria deberían volver a ser las mismas tras 1931. Los datos de 1931-1934 confirman esta hipótesis.

### *Marca la DD: una regresión sobre la depresión*

El cálculo DD más simple sólo involucra cuatro números, como en las ecuaciones (5.1) y (5.2). Pero en la práctica la receta DD se cocina mejor con modelos de regresión que se estiman por medio de muestras con más de cuatro puntos, como los 12 representados en la figura 5.2. Aparte de valorar más de dos periodos, la regresión DD incorpora con limpieza datos procedentes de más de dos entidades intercomparadas, como veremos en un análisis multiestatal de la MLDA en el apartado 5.2. No menos importante es el hecho de que la regresión DD facilita la inferencia estadística, un asunto que suele ser bastante delicado en la simple estimación por DD (véanse los detalles en el apéndice de este capítulo).

La receta de la regresión DD asociada a la figura 5.2 tiene tres ingredientes:

(i) Una variable binaria que identifica el distrito de tratamiento, que denominamos $TREAT_d$, donde el subíndice $d$ nos recuerda que el tratamiento difiere de un distrito a otro; $TREAT_d$ controla las diferencias fijas entre las unidades que se comparan.
(ii) Una variable binaria para los periodos postratamiento, que denominamos $POST_t$, donde el subíndice $t$ nos recuerda que varía

con el tiempo; $POST_t$ controla el hecho de que las condiciones varían con el tiempo para todo el mundo, se esté o no sometido a tratamiento.

(iii) El término de interacción $TREAT_d \times POST_t$, generado al multiplicar las dos variables binarias; el coeficiente de este término representa el efecto causal DD.

Consideramos que el tratamiento experimental de la era Caldwell es la provisión de crédito fácil en respuesta a una crisis de liquidez, de manera que $TREAT_d$ es igual a uno para los puntos del distrito sexto, y cero en los demás casos. La proporción de quiebras bancarias se moderó a partir de 1931, a medida que la crisis Caldwell remitía. Sin embargo, en la década de 1930 no quedaban bancos zombis: los bancos muertos desaparecían del todo. Las quiebras de la era Caldwell resultaron en un número menor de bancos abiertos en los años 1932-1934, a pesar de que la oficina de la Reserva Federal en San Luis hubiera empezado ya por entonces a prestar abiertamente. Por lo tanto, empleamos $POST_t$ para indicar todas las observaciones a partir de 1931. Para terminar, el término de interacción, $TREAT_d \times POST_t$, corresponde a observaciones del distrito sexto en el periodo posterior al tratamiento. Con más precisión, $TREAT_d \times POST_t$ corresponde a las observaciones del distrito sexto en los periodos en que la respuesta de la oficina de Atlanta a la crisis Caldwell fue relevante para los bancos que aún permanecían activos.

La regresión DD para el experimento Misisipi reúne todas estas piezas y estima:

$$Y_{dt} = \alpha + \beta\ TREAT_d + POST_t + \delta_{rDD}(TREAT_d \times POST_t) + e_{dt} \qquad (5.3)$$

en una muestra de tamaño 12. Esta muestra se construye a partir de las observaciones de ambos distritos y todos los años disponibles (seis años en cada distrito). El coeficiente del término de interacción, $\delta_{rDD}$, es el efecto causal de interés. Si se tienen sólo dos periodos, como en la figura 5.1., las estimaciones de $\delta_{DD}$ y $\delta_{rDD}$ coinciden (como consecuencia de las propiedades de la regresión con variables binarias descritas en el apéndice del capítulo 2). Con más de dos periodos, como en la figura 5.2, las estimaciones basadas en la ecuación (5.3)

deberían ser más precisas y tendrían que proporcionar una visión de los efectos de la política aplicada más fiable que la simple receta DD de cuatro números.[5]

Al ajustar la ecuación (5.3) a las 12 observaciones representadas en la figura 5.2 se generan las estimaciones siguientes (con errores típicos mostrados entre paréntesis):

$$\begin{aligned} Y_{dt} = 167 &- 29\ TREAT_d - 49\ POST_t \\ &+ 20.5\ (TREAT_d \times POST_t) + e_{dt}. \\ &(10.7) \end{aligned}$$

Estos resultados sugieren que unos 21 bancos se mantuvieron vivos gracias a los préstamos del distrito sexto. Esta estimación se parece a los 19 bancos que se salvaron a juzgar por el cálculo DD de cuatro números. El error típico del valor estimado para $\delta_{rDD}$ está en torno a 11, así que 21 constituye un resultado marginalmente significativo, lo mejor que cabría esperar a partir de una muestra tan pequeña.

### *Seamos realistas*

Parece muy probable que la oficina de la Reserva Federal en Atlanta salvara de la quiebra a muchos bancos del distrito sexto. Pero los bancos no tienen valor por sí mismos. ¿Favoreció la política de dinero fácil de la oficina de Atlanta la actividad económica real, es decir, a las empresas y empleos no bancarios? Las estadísticas de actividad empresarial dentro de los estados son escasas en este periodo. Aun así, los pocos números disponibles indican que la provisión de liquidez dada a los bancos desde Atlanta conllevó beneficios económicos. Esto se documenta en la tabla 5.1, que muestra los ingredientes para un análisis DD simple de los efectos de la liquidez aportada por la Reserva Federal sobre el número de empresas de ventas al mayor y sobre sus niveles de negocio.

Las estimaciones DD de las ventas al mayor en Misisipi son paralelas a las de los bancos. Entre 1929 y 1933 el número de empresas de

[5] De hecho, como explicamos en el apéndice de este capítulo, es difícil valorar la precisión de una estimación DD construida a partir de tan sólo dos unidades intercomparadas y dos periodos.

venta al mayor y sus volúmenes de negocio cayeron tanto en el distrito sexto como en el octavo, con un descenso mucho más brusco en este último, donde hubo más quiebras bancarias. En las décadas de 1920 y 1930 los negocios de venta al mayor dependían enormemente de los préstamos bancarios para financiar sus existencias. Las estimaciones de la tabla 5.1 sugieren que la reducción del crédito bancario en el distrito octavo en el arranque de la crisis Caldwell supuso también un freno en la actividad de venta al mayor, cuyos efectos probablemente se propagaron también a toda la economía local. Los vendedores al mayor del distrito sexto tuvieron más probabilidades de evitar este destino. Sin embargo, si se cocina con la receta simple DD de cuatro números, los indicios de los efectos del tratamiento de liquidez en la tabla 5.1 son más débiles que los que se deducen de la muestra, más amplia, sobre actividad bancaria.

El experimento Caldwell brinda una lección que salió bien cara sobre cómo atajar una crisis bancaria incipiente. Quizá el gobernador de la oficina de la Reserva Federal en San Luis, al ver que el colapso en el distrito sexto era más modesto que en el suyo, aprendió la

**Tabla 5.1. Empresas de venta al mayor: quiebras y ventas en 1929 y 1933**

| | 1929 | 1933 | Diferencia (1933-1929) |
|---|---|---|---|
| **Apartado A. Número de empresas de venta al mayor** | | | |
| Distrito sexto de la Reserva Federal (Atlanta) | 783 | 641 | –142 |
| Distrito octavo de la Reserva Federal (San Luis) | 930 | 607 | –323 |
| Diferencia (sexto-octavo) | –147 | 34 | 181 |
| **Apartado B. Ventas netas al mayor (millones de $)** | | | |
| Distrito sexto de la Reserva Federal (Atlanta) | 141 | 60 | –81 |
| Distrito octavo de la Reserva Federal (San Luis) | 245 | 83 | –162 |
| Diferencia (sexto-octavo) | –104 | –23 | 81 |

*Notas:* Esta tabla presenta un análisis DD de los efectos de la liquidez aportada por la Reserva Federal sobre el número de empresas de venta al mayor y el valor en dólares de sus ventas, para su comparación con el análisis DD de los efectos de la liquidez sobre la actividad bancaria de la figura 5.1.

lección y decidió cambiar de política en 1931. Pero las autoridades nacionales tardaron mucho más en comprender el poder paliativo de la política monetaria en una crisis financiera. Milton Friedman y su esposa Rose escribieron en sus memorias este famoso pasaje:

> En lugar de aplicar su poder para compensar la Depresión, [el comité directivo de la Reserva Federal en Washington D.C.] decidió una reducción de un tercio en la cantidad de dinero entre 1929 y 1933. Si se hubiera comportado como pretendían sus fundadores, habría evitado tal declive y, de hecho, lo habría convertido en el aumento que se necesitaba para propiciar el crecimiento normal de la economía.[6]

Lo que no resulta sencillo es decir si el problema del manejo de las crisis financieras se ha resuelto desde entonces. Los complejos mercados financieros actuales descarrilan por muchos motivos, y no todos ellos los puede contener la Reserva Federal imprimien-

[6] Milton Friedman y Rose D. Friedman, *Two Lucky People: Memoirs*, University of Chicago Press, 1988, página 233.

do dinero. Las autoridades monetarias de nuestros tiempos están aprendiendo esa dura lección.

## 5.2 Bebe y vive

Shen: ¿Estás dispuesto a morir por la verdad?
Po: ¡Ya lo creo!... Aunque prefiero no hacerlo.

*Kung Fu Panda 2*

Tras la retirada de la prohibición federal del consumo de alcohol en 1933, cada estado de Estados Unidos tuvo libertad para regular esta materia. La mayoría instituyó una edad mínima legal para el consumo de alcohol (MLDA, «mínimum legal drinking age») de 21 años, pero Kansas, Nueva York y Carolina del Norte permitieron beber a los 18. Tras la vigésimo sexta enmienda a la constitución de 1971, que redujo a los 18 años la edad para tener derecho al voto en respuesta a las turbulencias desatadas por la guerra de Vietnam, muchos estados redujeron a su vez la MLDA. Pero no todos: Arkansas, California y Pensilvania se cuentan entre los estados que mantienen el límite en 21 años. En 1984, el acta nacional sobre edad mínima para el consumo de alcohol castigó el consumo de alcohol entre los jóvenes al denegar las ayudas federales para la construcción de autopistas a los estados que situaban la MLDA en 18 años. En 1988, la totalidad de los 50 estados y el distrito de Columbia habían optado ya por fijar la MLDA en 21 años, aunque algunos habían captado el mensaje de las autovías federales más rápido que otros.

Como sucede con muchas políticas en Estados Unidos, la interacción entre la legislación federal y la de los estados genera un tejido multicolor y cambiante de normas legales. Esta variación política es un vergel para los maestros de la econometría: la variación de las leyes sobre MLDA en los estados es fácil de explotar en el contexto DD. Este marco proporciona una alternativa al enfoque RD del capítulo 4 en nuestros esfuerzos por desvelar los efectos de las políticas sobre el consumo de alcohol.[7]

[7] Carpenter y Dobkin, en «The Minimum Legal Drinking Age», *Journal of Economic Perspectives*, 2011, analizaron la MLDA mediante DD.

### *Patrones a partir de retales*

Alabama redujo la MLDA a 19 años en 1975, pero el estado de Arkansas, cercano tanto en lo alfabético como en lo geográfico, mantuvo el umbral en 21 desde el fin de la prohibición. La indulgencia de Alabama para con sus bebedores jóvenes, ¿le costó la vida a alguno de ellos? Tratamos esta cuestión estimando un modelo de regresión DD para los datos de mortalidad en las edades de 18 a 20 años entre 1970 y 1983. La variable dependiente se representa como $Y_{st}$ y mide las tasas de mortalidad en el estado $s$ y el año $t$. Con una muestra que incluye sólo Alabama y Arkansas, el modelo de regresión DD para $Y_{st}$ toma la forma

$$\begin{aligned} Y_{st} = {} & \alpha + \beta TREAT_s + \gamma\, POST_t \\ & + \delta_{rDD}(TREAT_s \times POST_t) + e_{st}, \end{aligned} \tag{5.4}$$

donde $TREAT_s$ es una variable binaria que identifica el estado de Alabama, $POST_t$ es otra binaria que indica los años desde 1975 y posteriores, y el término de interacción $TREAT_s \times POST_t$ indica las observaciones para Alabama durante los años en que se permitía beber con edades más tempranas. El coeficiente $\delta_{rDD}$ refleja el efecto de situar la MLDA en 19 años sobre las tasas de mortalidad.

La ecuación (5.4) es comparable al modelo de regresión DD para los dos distritos de la Reserva Federal en Misisipi. Pero ¿por qué limitar el estudio a Alabama y Arkansas? Los registros legislativos permiten más de un experimento sobre la MLDA. Por ejemplo, la MLDA cayó en Tennessee hasta los 18 años en 1971, para luego subir a 19 en 1979. Una consecuencia complicada pero manejable de estas diferencias temporales en las reducciones de la MLDA en Alabama y Tennessee es la ausencia de un periodo común posterior al tratamiento. Cuando se combinan múltiples experimentos sobre MLDA en un esquema DD, se pasa de tener una única variable binaria $POST_t$ a tener todo un conjunto de ellas que indican cada año de la muestra, excepto uno que se mantiene como referencia. Los coeficientes de estas variables binarias, conocidas como *efectos temporales*, reflejan los cambios de los índices de mortalidad comunes a todos los estados con el paso del tiempo.[8]

[8] Incluimos en los datos un efecto temporal menos que el número de años. Los efectos temporales miden los cambios en función del tiempo respecto de un punto de partida, normalmente el primer año de la muestra.

Nuestro método de regresión DD multi-MLDA debería reflejar también el hecho de que hay muchas comparaciones de tipo causal, debidas a los cambios ocurridos en muchos estados. En lugar de controlar sólo la diferencia entre, digamos, los distritos sexto y octavo de la Reserva Federal, como en el experimento Misisipi del apartado 5.1, o la diferencia entre Alabama y Arkansas en el ejemplo anterior, el procedimiento seguido con muchos estados controla por las distintas tasas de mortalidad en cada uno de estos estados. Esto se logra mediante la introducción de unos *efectos estatales* como un conjunto de variables binarias para cada estado de la muestra, excepto para uno, que se omite como grupo de referencia. Una regresión DD de los datos de Alabama, Arkansas y Tennessee, por ejemplo, incluiría dos efectos estatales. Los efectos estatales sustituyen a la única variable binaria $TREAT_s$ del análisis comparativo entre sólo dos estados (dos grupos).

En este contexto surge una dificultad adicional debida a la ausencia de una variable de tratamiento común que se active o desactive de manera discreta. La MLDA adopta valores entre 18 y 21, lo que genera efectos del tratamiento «consumo legal de alcohol» para las edades de 18, 19 o 20 años. Los maestros de la econometría simplifican estas situaciones reduciéndolas a una medida única de la exposición a la política de interés, en este caso el acceso al alcohol. Nuestra estrategia de simplificación sustituye $TREAT_d \times POST_t$ por una variable que denotaremos como $LEGAL_{st}$. Esta variable mide la proporción de personas entre 18 y 20 años que legalmente pueden consumir alcohol en el estado $s$ y el año $t$. En algunos estados no se permite beber a nadie por debajo de 21 años de edad, mientras que hay otros en los que la MLDA es igual a 19, es decir, aproximadamente un tercio de las personas entre 18 y 21 puede beber, y en los estados con MLDA igual a 18 todas las personas de 18 a 21 lo tienen permitido. Nuestra definición de $LEGAL_{st}$ también capta la variación debida a cambios en el curso de los años. Por ejemplo, la MLDA pasó a valer 19 en Alabama en julio de 1975. En consecuencia, $LEGAL_{\text{AL,1975}}$ se escala para reflejar el hecho de que las personas de 19 a 20 años en Alabama pudieron beber legalmente durante sólo la mitad de ese año.

El modelo de regresión DD multiestatal es:

$$Y_{st} = \alpha + \delta_{rDD} LEGAL_{st} + \sum_{k=\text{Alaska}}^{\text{Wyomig}} \beta_k STATE_{ks} + \sum_{j=1971}^{1983} \gamma_j YEAR_{jt} + e_{st}. \quad (5.5)$$

No hay que dejarse intimidar por los grandes sumatorios de esta ecuación. Esta notación describe de manera compacta modelos con muchas variables binarias, como sucedía con los modelos con variables binarias para grupos selectivos en el capítulo 2. Aquí todos los estados salvo uno (el de referencia) tienen su propia variable binaria, indexada con el subíndice $k$ para cada caso. El índice $s$ sigue la pista de qué estado es el que aporta las observaciones. La variable binaria estatal $k$-ésima, $STATE_{ks}$, vale la unidad cuando una observación procede del estado $k$, lo que implica que $s = k$, y vale cero en otro caso. Las observaciones de California, por ejemplo, tienen activada $STATE_{\text{CA},s}$, y desactivadas todas las demás binarias estatales.

Los efectos estatales, $\beta_k$, son los coeficientes de las variables binarias estatales. Por ejemplo, el efecto estatal de California, $\beta_{\text{CA}}$, es el coeficiente de $STATE_{\text{CA},s}$. Cada estado, salvo el de referencia (el cual se omite al construir las variables binarias estatales) tiene un efecto estatal en la ecuación (5.5). Como hay tantas, usamos una notación con sumatorios, $\sum_{k=\text{Alaska}}^{\text{Wyoming}} \beta_k STATE_{ks}$, para no tener que escribirlas todas. Los efectos temporales, $\gamma_t$, corresponden igualmente a los coeficientes de las variables binarias de los años, $YEAR_{jt}$. Éstas se activan cuando las observaciones de los datos son del año $j$, es decir, cuando $t = j$. Por eso las llamamos también *efectos anuales*. El efecto anual de 1975, $\gamma_{1975}$, es el coeficiente de $YEAR_{1975,t}$. Aquí, de nuevo, cada año de la muestra salvo uno, que sirve de referencia, tiene su efecto anual, y por eso se emplean sumatorios en la notación para escribirlo todo de manera más compacta.[9]

[9] He aquí otro modo de ver cómo funciona la notación. Consideremos una observación para $s$ = NY (Nueva York). Entonces tenemos:

$$\sum_{k=\text{Alaska}}^{\text{Wyoming}} \beta_k STATE_{ks} = \beta_{\text{NY.}}$$

de modo que la suma de todas las posibles variables binarias estatales capta el efecto de Nueva York, $\beta_{\text{NY}}$, cuando las observaciones son de Nueva York. Todas las demás variables binarias de la suma son cero. En consecuencia, si $t$ = 1980, entonces tenemos:

$$\sum_{j=1971}^{1983} \gamma_j YEAR_{jt} = \gamma_{\text{NY.}}$$

y la suma recoge el efecto anual de 1980 cuando las observaciones son de ese año.

Nuestro análisis MLDA multiestatal usa un conjunto de datos que abarca 14 años y 51 estados (incluido el distrito de Columbia), con un total de 714 observaciones. Esta estructura de datos se denomina *tabla estado-año*. Los efectos estatales en la ecuación (5.5) controlan diferencias fijas entre estados (por ejemplo, los accidentes de tráfico mortales son más frecuentes, en promedio, en los estados rurales con velocidades de circulación medias más elevadas). Los efectos temporales (anuales) de la misma ecuación controlan tendencias en mortalidad que son comunes a todos los estados (debidos, por ejemplo, a tendencias nacionales en hábitos de bebida o en seguridad de los vehículos). La ecuación (5.5) atribuye los cambios en mortalidad dentro de los estados a cambios en la variable $LEGAL_{st}$. Como veremos a continuación, esta atribución causal constituye un supuesto de existencia de una tendencia común, como en nuestro análisis de las quiebras bancarias inducidas por la crisis de Caldwell en el apartado anterior.

Las estimaciones de $\delta_{rDD}$ que brinda la ecuación (5.5) sugieren que el acceso legal al alcohol causó unas 11 muertes adicionales por cada 100.000 personas entre 18 y 20 años de edad, de las cuales hay entre siete y ocho muertes debidas a accidentes de tráfico. Estos resultados constan en la primera columna de la tabla 5.2 y son algo mayores que las estimaciones RD aportadas en la tabla 4.1 (capítulo 4), pero aun así son coherentes con ellas a grandes rasgos. Las estimaciones de accidentes de tráfico de la tabla 5.2 también son razonablemente precisas, con errores típicos en torno a 2,5. Hay que remarcar que, al igual que en las estimaciones por medio de RD, este modelo de regresión DD aporta pocos indicios a favor de que haya un efecto del consumo legal de alcohol sobre la mortalidad. Los indicios que da la regresión DD acerca de posibles efectos sobre laa tasa de suicidios son más débiles que los que se deducen mediante RD en la tabla 4.1. Al mismo tiempo, ambas estrategias coinciden en que el incremento del número de suicidios es inferior que el de las muertes por accidente de tráfico.

### *Pongamos a prueba las suposiciones DD*

Las muestras que incluyen muchos estados y años nos permiten relajar el supuesto de que exista una tendencia común, es decir, podemos introducir un cierto grado de evolución distinta en los resultados

**Tabla 5.2. Estimaciones mediante regresión DD de los efectos de la MLDA sobre las tasas de mortalidad**

| Variable dependiente | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Todas las muertes | 10.80 | 8.47 | 12.41 | 9.65 |
| | (4.59) | (5.10) | (4.60) | (4.64) |
| Accidentes de tráfico | 7.59 | 6.64 | 7.50 | 6.46 |
| | (2.50) | (2.66) | (2.27) | (2.24) |
| Suicidio | .59 | .47 | 1.49 | 1.26 |
| | (.59) | (.79) | (.88) | (.89) |
| Todas las causas internas | 1.33 | .08 | 1.89 | 1.28 |
| | (1.59) | (1.93) | (1.78) | (1.45) |
| Tendencias estatales | No | Sí | No | Sí |
| Pesos | No | No | Sí | Sí |

*Notas:* Esta tabla muestra las estimaciones mediante regresión DD de los efectos de la edad mínima legal para el consumo de alcohol (MLDA) sobre las tasas de mortalidad (cada 100.000 personas) para edades entre 18 y 20 años. La tabla muestra los coeficientes para la proporción de bebedores legales por estado y año a partir de modelos que incluyen controles para los efectos estatales y anuales. Los modelos usados para construir las estimaciones de las columnas (2) y (4) incluyen tendencias temporales lineales específicas para cada estado. Las columnas (3) y (4) muestran estimaciones mediante mínimos cuadrados ponderados, usando la población de cada estado como peso. El tamaño de la muestra es 714. Se dan los errores típicos entre paréntesis

de un estado a otro, cuando no hay efecto tratamiento. Un modelo de regresión DD que controla por las tendencias propias de cada estado tiene este aspecto:

$$\begin{aligned} Y_{st} = {} & \alpha + \delta_{rDD} LEGAL_{st} \\ & + \sum_{k=\text{Alaska}}^{\text{Wyoming}} \beta_k LEGAL_{ks} + \sum \gamma_j YEAR_{jt} \\ & + \sum_{k=\text{Alaska}}^{\text{Wyoming}} \delta_k (STATE_{ks} \times t)_{ks} + \sum \gamma_j YEAR_{jt} \end{aligned} \tag{5.6}$$

Este modelo presupone que, en ausencia de un efecto tratamiento, la mortalidad en el estado $k$ se desvía de los efectos anuales comunes según una tendencia lineal representada por el coeficiente $\theta_k$.

Hasta ahora y por ahora hemos insistido en que la clave del método DD está en las tendencias comunes. ¿Cómo es posible, entonces, que ahora consideremos modelos como el de la ecuación (5.6), que relajan este supuesto central de la existencia de una tendencia común a todos los estados? Para ver el funcionamiento de tales modelos consideremos una muestra de dos estados: el primero, Alamar, redujo la MLDA a 18 en 1975, mientras que el estado vecino de Alabastro la mantuvo en 21. Como punto de partida, la figura 5.4 traza la evolución de la tendencia común en ambos estados. Las muertes por cada 100.000 habitantes evolucionan de manera paralela hasta 1975 (casi todo empeoró en los años setenta, por eso representamos un incremento de la mortalidad). Además, las tasas de mortalidad dan un salto por encima de la tendencia común en Alamar en 1975, cuando ese estado redujo su MLDA. Dado el paralelismo y las fechas, parece justo culpar de ese salto a la bajada de la MLDA en Alamar.

La figura 5.5 traza un escenario con una tendencia más elevada en Alamar que en Alabastro. Como sucedía en los datos de la figura anterior, una regresión simple DD aplicada a este caso generaría unas estimaciones que nos harían culpar a la MLDA (el contraste «post menos pre» en Alamar es mayor que en Alabastro). Pero en este caso la estimación DD resultante sería espuria: la diferencia entre las tendencias estatales es previa a la liberalización de la MLDA en Alamar y, por tanto, no puede guardar relación con la misma.

Por fortuna, tales diferencias de tendencia se pueden representar por medio de los parámetros de tendencia propia de cada estado, $\theta_k$, de la ecuación (5.6). En los modelos que incluyen controles de las tendencias estatales específicas, los efectos de la MLDA se reflejan en desviaciones bruscas sobre tendencias que por lo demás son suaves, incluso cuando esas tendencias no sean comunes. La figura 5.6 muestra cómo una regresión DD capta los efectos del tratamiento superpuestos a tendencias no comunes. La mortalidad en Alamar crece a un ritmo mayor que en Alabastro a lo largo del mismo periodo. Pero el incremento en Alamar resulta especialmente pronunciado entre 1974 y 1975, cuando este estado rebajó su MLDA. El coeficiente $LEGAL_{st}$ de la ecuación (5.6) capta este hecho, a la vez que el modelo permite encajar el hecho de que las tasas de mortalidad en estados distintos siguieran trayectorias diferentes ya desde el comienzo.

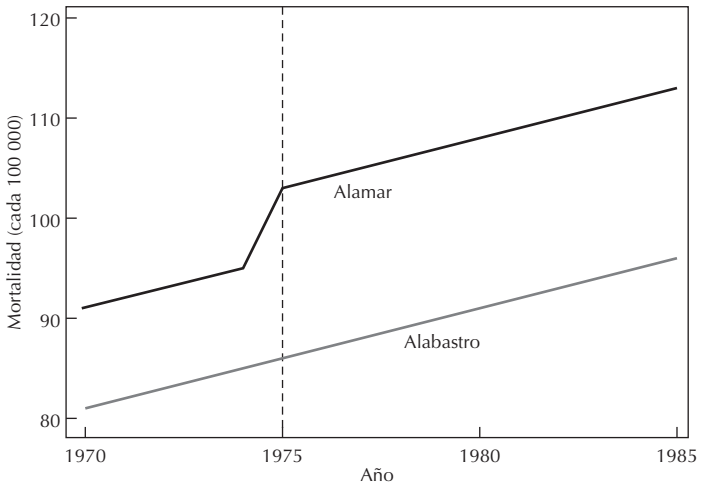



**Figura 5.4.** Un efecto MLDA en estados con tendencias paralelas.

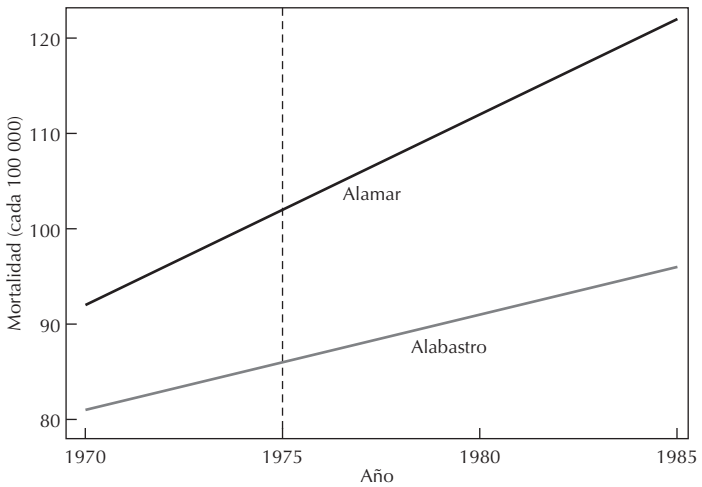



**Figura 5.5.** Un efecto MLDA espurio en estados con tendencias no paralelas.

Los modelos con tendencias estatales específicas lineales permiten comprobar la interpretación causal de cualquier conjunto de estimaciones mediante regresión DD usando datos de múltiples periodos. Pero en la práctica la realidad empírica puede ser mucho más difusa y difícil de interpretar que los ejemplos idealizados que se muestran en las figuras 5.4 a 5.6. Los resultados generados por un modelo de

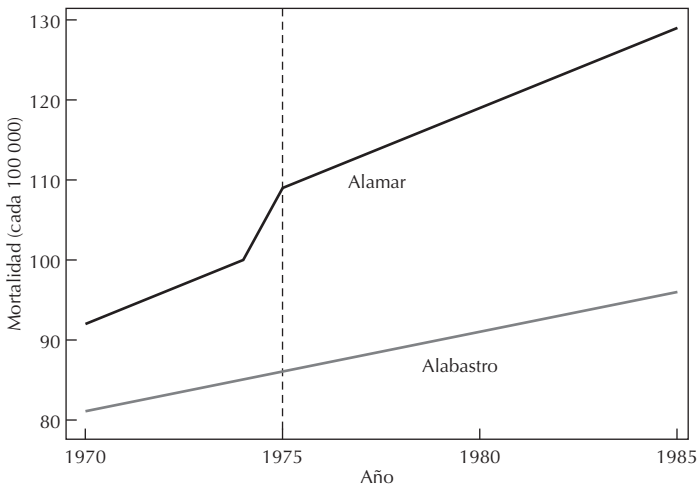



**Figura 5.6.** Un efecto MLDA real, visible aunque las tendencias no sean paralelas.

regresión como el de la ecuación (5.6) suelen ser imprecisos. Cuanto más brusca sea la desviación respecto de la tendencia debida a un efecto causal, más probable será que logremos descubrirla. Por otra parte, si los efectos del tratamiento se ponen de manifiesto solamente de manera gradual, entonces las estimaciones de ecuaciones como la (5.6) pueden fracasar en el intento de distinguir entre los efectos del tratamiento y las tendencias diferenciales, con un resultado final impreciso y, por tanto, no concluyente.

Por fortuna, en el análisis causal de los efectos de la MLDA por medio de DD, la introducción de tendencias estatales específicas afecta muy poco a las estimaciones de la regresión DD. Esto se aprecia en la columna (2) de la tabla 5.2, que recoge las estimaciones mediante regresión DD del efecto de la MLDA con el modelo de la ecuación (5.6). Al añadir las tendencias diferentes se incrementan un poco los errores típicos, pero la pérdida de precisión en este caso es modesta. Los resultados de la columna (2) contribuyen a confirmar una interpretación causal de los efectos MLDA más precisos que aparecen en la columna (1) de la tabla.

Los procesos de decisión de los estados constituyen un asunto complejo, con cambios frecuentes en muchos frentes. Las estimaciones DD de los efectos de la MLDA, con o sin tendencias estatales específicas, podrían estar sesgadas por cambios legislativos contem-

poráneos en otras áreas. Una consideración importante cuando se trata de alcohol, por ejemplo, es la de su precio. Los impuestos son la herramienta más poderosa que puede usar el gobierno para alterar el precio de tu bebida favorita. Muchos estados imponen unas tasas elevadas a la cerveza, que puede evaluarse en dólares por cada unidad de volumen de alcohol, por galón en este caso.[10] Los impuestos sobre la cerveza van desde apenas unos centavos por galón hasta más de un dólar por galón en algunos estados del sur. Estos impuestos cambian de vez en cuando, normalmente al alza, para desesperación del Instituto de la Cerveza (con una tasa impositiva de 2 centavos por galón desde 1935, Wyoming es el paraíso de la cerveza). Puede tener sentido que los estados quieran incrementar los impuestos sobre la cerveza a la vez que aumentan su MLDA, quizá en el marco de un esfuerzo más amplio para reducir el consumo de alcohol. Si esto es así, deberíamos controlar por los cambios en estos impuestos a la hora de estimar los efectos de la MLDA.

Los modelos de regresión DD que incluyen controles por los impuestos sobre la cerveza conducen a estimaciones similares a las que resultan cuando no se introducen tales controles. Esto se puede comprobar en la tabla 5.3, que refleja los coeficientes estimados para $LEGAL_{st}$ y para los impuestos estatales sobre la cerveza en modelos para las cuatro tasas de mortalidad examinadas en la tabla 5.2. Las columnas (1) y (2) de la tabla 5.3 muestran los efectos de los impuestos y de la MLDA estimados usando una regresión simple, sin controles para tendencias estatales específicas, mientras que las columnas (3) y (4) proceden de otra regresión que incluye tales controles de tendencia. Los efectos de los impuestos se estiman con menos precisión que los de la MLDA, probablemente porque los impuestos sobre la cerveza cambian con menos frecuencia que la MLDA. Las estimaciones de los impuestos sobre la cerveza a partir de modelos que incluyen tendencias estatales propias resultan especialmente ruidosas. Aun así, el Instituto de la Cerveza estará encantado de saber que estos resultados no favorecen que se incrementen más los impuestos sobre esta bebida. Nos agrada también constatar que nuestras estimaciones MLDA son robustas a la incorporación de controles para los impuestos sobre la cerveza. ¡Brindaremos con una cerveza para celebrarlo!

[10] El galón estadounidense equivale a 3,79 litros. (*N. de la T.*)

**Tabla 5.3. Estimaciones mediante regresión DD de los efectos de la MLDA, con control de los impuestos sobre la cerveza**

| | Sin tendencias | | Con tendencias | |
|---|---|---|---|---|
| | Fracción legal | Impuesto sobre la cerveza | Fracción legal | Impuesto sobre la cerveza |
| Variable dependiente | (1) | (2) | (3) | (4) |
| Todas las muertes | 10.98 | 1.51 | 10.03 | –5.52 |
| | (4.69) | (9.07) | (4.92) | (32.24) |
| Accidentes de tráfico | 7.59 | 3.82 | 6.89 | 26.88 |
| | (2.59) | (5.40) | (2.66) | (20.12) |
| Suicidio | .45 | –3.05 | .38 | –12.13 |
| | (.60) | (1.63) | (.77) | (8.82) |
| Causas internas | 1.46 | –1.36 | .88 | –10.31 |
| | (1.61) | (3.07) | (1.81) | (11.64) |

*Notas:* Esta tabla muestra las estimaciones mediante regresión DD de los efectos de la edad mínima legal para el consumo de alcohol (MLDA) sobre las tasas de mortalidad (cada 100.000 personas) para edades entre 18 y 20 años, incluyendo controles para los impuestos estatales sobre la cerveza. La tabla indica los coeficientes para la proporción de bebedores legales por estado y año, y para los impuestos sobre la cerveza por estado y año, a partir de modelos que incluyen controles para los efectos estatales y anuales. Las variables de fracción legal y de impuestos sobre la cerveza se incluyen en un modelo de regresión simple, estimado sin tendencias, y que conduce a las estimaciones de las columnas (1) y (2), así como en otro modelo con tendencias estatales lineales específicas que conduce a los valores de las columnas (3) y (4). El tamaño de la muestra es 700. Los errores típicos constan entre paréntesis.

### *¿Qué estás pesando?*

Las estimaciones procedentes de las ecuaciones (5.5) y (5.6) que constan en las columnas (1) y (2) de la tabla 5.2 dan el mismo peso a todas las observaciones, como si los datos procedentes de cada estado fueran igual de valiosos. Pero los estados no se crearon iguales, al menos en un aspecto importante: algunos, como Tejas o California, son más grandes que la mayoría de países del mundo, mientras que otros, como Vermont o Wyoming, tienen poblaciones inferiores a las de muchas ciudades de Estados Unidos. Podríamos preferir estimaciones que reflejasen este hecho dando más peso a los estados más poblados. El procedimiento de regresión que permite hacer esto se denomina *mínimos cuadrados ponderados* (MCP). El estimador

estándar MCO estima una recta minimizando el promedio muestral de residuos cuadráticos, de manera que cada residuo cuadrático entra con el mismo peso en la suma.[11] Tal como sugiere su nombre, los MCP dan un peso a cada término en la suma de residuos cuadráticos, según la población o según cualquier otro criterio elegido por el investigador.

La ponderación por población tiene dos consecuencias. Primero, como se indicó en el capítulo 2, los modelos de regresión de los efectos de los tratamientos representan un promedio ponderado de los efectos sobre los grupos o células representados en los datos. En una tabla estado-año esos grupos son los estados. El cálculo por MCO produce estimaciones de efectos causales promedio que ignoran el tamaño de las poblaciones, de manera que los resultados corresponden a promedios entre estados, no entre individuos. La asignación de pesos según la población genera un promedio ponderado en el que los efectos causales de estados como Tejas ejercen más influencia que los de estados como Vermont. La asignación de pesos según la población puede parecer atractiva, pero quizá no lo sea tanto. Es cierto que el ciudadano medio tiene más posibilidades de vivir en Tejas que en Vermont, pero los cambios en la MLDA de Vermont proporcionan una variación que puede ser igual de útil que los cambios en Tejas. Cabría esperar, por tanto, que las estimaciones por regresión de una tabla estado-año no sean demasiado sensibles a esta asignación de pesos.

La ponderación según población pueden incrementar la precisión de las cantidades estimadas. Como en Vermont hay muchísimos menos conductores que en Tejas, es de esperar que las tasas de mortalidad por accidentes de tráfico en Vermont cambien más de un año a otro que las de Tejas (esto refleja la variación muestral tratada en el apéndice del capítulo 1). En un sentido estadístico, los datos de Tejas son más fiables y, por tanto, quizá merezcan tener más peso. Pero de nuevo sucede que la cosa no está tan clara. Desde el punto de vista de la teoría econométrica, los maestros pueden afirmar que las estimaciones ponderadas son más precisas que las no ponderadas

[11] Los residuos de la regresión, definidos en el apéndice del capítulo 2, son las diferencias entre los valores ajustados generados por el modelo que se está estimando y la variable dependiente del modelo.

sólo si se verifica una serie de condiciones técnicas restrictivas.[12] De nuevo nos encontramos con que el mejor escenario es aquel que nos ofrece un conjunto de resultados (es decir, estimaciones y sus errores típicos) razonablemente insensibles a las ponderaciones.

Las columnas (3) y (4) de la tabla 5.2 contienen estimaciones mediante MCP de las ecuaciones (5.5) y (5.6). Se corresponden con las estimaciones por MCO de las columnas (1) y (2) de la misma tabla, pero el estimador por MCP se pondera cada observación según la población del estado con edades entre 18 y 20 años. Por fortuna para nuestra comprensión de los efectos de la MLDA, la introducción de ponderaciones aquí importa poco. Podría parecer de nuevo que los maestros abstemios son recompensados por su virtud.

Maestro Stevefu: Hazme un resumen, Pequeño Saltamontes.

Pequeño Saltamontes: Los grupos de tratamiento y de control pueden diferir en ausencia de tratamiento y, sin embargo, moverse en paralelo. Este hecho abre la puerta a la estimación de efectos causales con la técnica DD.

Maestro Stevefu: ¿Por qué la técnica DD es mejor que la simple comparación entre dos grupos?

Pequeño Saltamontes: Al comparar cambios, en vez de niveles, eliminamos las diferencias fijas entre grupos que de otro modo podrían generar un sesgo de variables omitidas.

Maestro Stevefu: ¿Cómo se aplica la técnica DD con múltiples grupos de comparación y para múltiples años?

Pequeño Saltamontes: He observado el poder y la flexibilidad de la regresión DD, Maestro. Por ejemplo, en una tabla estado-año con políticas de los estados que cambian con el tiempo, como la MLDA, sólo necesitamos introducir controles de carácter estatal y anual.

Maestro Stevefu: ¿De qué depende la ventura de las estimaciones DD?

[12] Uno de los requisitos es que la función de valor esperado condicionado (FVEC) subyacente sea lineal. El apéndice del capítulo 2 aclara, sin embargo, que muchos modelos de regresión son sólo aproximaciones lineales a la verdadera FVEC.

Pequeño Saltamontes: De las tendencias paralelas, del supuesto de que, en ausencia de tratamiento, los resultados de los grupos de tratamiento y de control se moverían en paralelo. La técnica DD vive y muere por esta condición. Aunque se pueden permitir tendencias estatales específicas lineales cuando la tabla es lo bastante grande, los maestros confían en que los resultados no cambien con su inclusión.

**Maestros de la econometría: John Snow**

El médico británico John Snow fue uno de los padres de la epidemiología moderna, el estudio de la evolución de las enfermedades en una población. Con su estudio de la epidemia de cólera en Londres en 1849 desafió la idea convencional de que la enfermedad la causaba un aire putrido. Creyó que el cólera podría estar causado por agua insalubre, una idea que plasmó por primera vez en su ensayo de 1849 *On the Mode of Communication of Cholera* [«Sobre el modo de contagio del cólera»].

Otra epidemia de cólera en 1853 y 1854 segó muchas vidas en el barrio londinense del Soho. Snow atribuyó la epidemia del Soho al agua de una fuente situada en Broad Street. Sin miedo a dar una ayudita a un experimento natural, consiguió convencer a la autoridad local de que retirara el mango que permitía bombear el agua. Las muertes por cólera en el Soho remitieron poco después, pero Snow reparó en que los índices de mortalidad en su zona de tratamiento de Broad Street habían empezado a descender antes, lo que hacía difícil interpretar los datos de su experimento natural. La técnica DD era tan tornadiza cuando nació como lo es ahora.

Snow era un triturador de datos meticuloso y dejó el listón a una altura que aún aspiramos a alcanzar. En una revisión de su ensayo en 1855 Snow hizo constar las muertes por distritos y por fuentes de agua en varias partes de Londres. Se percató de que muchos de los distritos con índices de mortalidad elevados en el sur de Londres recibían agua de dos empresas suministradoras, o bien la Southwalk and Vauxhall, o bien la Lambeth. En 1849 ambas compañías extraían el agua del contaminado río Támesis a su paso por el centro de Londres. Pero a partir de 1852 la compañía Lambeth empezó a traerla de

Thames Ditton, un manantial no contaminado situado en un tramo más alto del río. Snow mostró que entre 1849 y 1854 las muertes por cólera disminuyeron en la zona abastecida por Lambeth, pero aumentaron en la atendida por Southwalk and Vauxhall. Nuestra figura 5.7 reproduce la tabla 12 del ensayo de Snow de 1855.[13] Esta tabla contiene los ingredientes del análisis DD en dos periodos que efectuó Snow sobre los índices de mortalidad en función del origen del agua.

## Apéndice: Errores típicos en regresiones DD

La regresión DD es un caso especial de estimación a partir de tablas de datos. Una tabla estado-año consiste en observaciones repetidas en varios estados a lo largo del tiempo. La estructura repetitiva de tales conjuntos de datos suscita problemas estadísticos especiales. Los datos económicos de este tipo suelen mostrar una propiedad llamada *autocorrelación*. Los datos con autocorrelación presentan el rasgo persistente de que es probable que los valores de las variables para periodos cercanos entre sí se parezcan.

Esperamos autocorrelación en series de datos temporales como las tasas de desempleo anuales. Cuando la tasa de desempleo de un estado es un año superior a la media, es probable que también lo sea en el siguiente. Como los conjuntos de datos tabulados combinan observaciones repetidas para cada estado individual (en nuestro ejemplo sobre la MLDA) o para regiones concretas (en nuestro experimento Misisipi), tales datos suelen tener autocorrelación. Cuando la variable dependiente de una regresión está afectada de autocorrelación, los residuos de cualquier modelo de regresión que pretenda explicar esta variable suelen estar también autocorrelacionados. Cuando se combinan residuos con autocorrelación y regresores con autocorrelación, hay que cambiar la fórmula para calcular los errores típicos.

Si ignoramos la autocorrelación y recurrimos a la fórmula simple de los errores típicos, la ecuación (2.15), es probable que las conclusiones estadísticas a las que lleguemos estén equivocadas. La penitencia por ignorar la autocorrelación consiste en exagerar la precisión

[13] John Snow, *On the Mode of Communication of Cholera*, John Churchill, segunda edición, 1855.

TABLE XII.

| Sub-Districts. | Deaths from Cholera in 1849. | Deaths from Cholera in 1854. | Water Supply. |
|---|---|---|---|
| St. Saviour, Southwark | 283 | 371 | |
| St. Olave | 157 | 161 | |
| St. John, Horsleydown | 192 | 148 | |
| St. James, Bermondsey | 249 | 362 | |
| St. Mary Magdalen | 259 | 244 | |
| Leather Market | 226 | 237 | |
| Rotherhithe* | 352 | 282 | Southwark & Vauxhall Company only. |
| Wandsworth | 97 | 59 | |
| Battersea | 111 | 171 | |
| Putney | 8 | 9 | |
| Camberwell | 235 | 240 | |
| Peckham | 92 | 174 | |
| Christchurch, Southwark | 256 | 113 | |
| Kent Road | 267 | 174 | |
| Borough Road | 312 | 270 | |
| London Road | 257 | 93 | |
| Trinity, Newington | 318 | 210 | |
| St. Peter, Walworth | 446 | 388 | Lambeth Company, and Southwark and Vauxhall Compy. |
| St. Mary, Newington | 143 | 92 | |
| Waterloo Road (1st) | 193 | 58 | |
| Waterloo Road (2nd) | 243 | 117 | |
| Lambeth Church (1st) | 215 | 49 | |
| Lambeth Church (2nd) | 544 | 193 | |
| Kennington (1st) | 187 | 303 | |
| Kennington (2nd) | 153 | 142 | |
| Brixton | 81 | 48 | |
| Clapham | 114 | 165 | |
| St. George, Camberwell | 176 | 132 | |
| Norwood | 2 | 10 | |
| Streatham | 154 | 15 | Lambeth Company only. |
| Dulwich | 1 | — | |
| Sydenham | 5 | 12 | |
| First 12 sub-districts | 2261 | 2458 | Southwk. & Vauxhall. |
| Next 16 sub-districts | 3905 | 2547 | Both Companies. |
| Last 4 sub-districts | 162 | 37 | Lambeth Company. |

* A small part of Rotherhithe is now supplied by the Kent Water Company.

**Figura 5.7.** La receta DD de John Snow.

de las estimaciones de la regresión. Esto es así porque la teoría muestral de la inferencia mediante regresión que planteamos en el apéndice del capítulo 1 presupone que los datos disponibles proceden de muestras aleatorias. La autocorrelación constituye una desviación de la aleatoriedad, con la consecuencia crucial de que cada nueva observación en una serie temporal con autocorrelación contiene menos información que si procediera de una muestra aleatoria.

Así como los errores típicos robustos tratados en el apéndice del capítulo 1 corrigen el carácter heterocedástico, hay una fórmula modificada para los errores típicos que da respuesta al desafío de la autocorrelación. La fórmula apropiada en este caso se denomina *error típico robusto por grupos* (*clustered standard error*). La expresión para el error típico robusto por grupos es más compleja que la de los errores robustos dada en la ecuación (2.16), y no hay que aprendérsela para el examen. Lo importante es que el trabajo con grupos (una opción disponible en la mayoría de los paquetes informáticos de regresión) permite trabajar con datos en los que hay autocorrelación dentro de los grupos definidos por el investigador. En contraste con el supuesto de que los datos están muestreados al azar, la fórmula de los errores típicos robustos por grupos requiere muestrear al azar los grupos, sin formular ningún supuesto restrictivo acerca de lo que contengan en su interior.

En el ejemplo MLDA tratado en este capítulo, los grupos son los estados. Pero es frecuente que lo que aparezca de manera repetida en nuestros datos sean personas individuales. Los participantes en el estudio RAND HIE aportaron hasta cinco observaciones anuales sobre su uso de los servicios de salud en la muestra empleada para construir la tabla 1.4, y los estudiantes aparecen en dos cursos distintos en la muestra empleada para estimar el modelo del efecto de la calidad de los demás estudiantes, ecuación (4.9). En estos ejemplos se corrige el hecho de que los diversos resultados procedentes de una misma persona tienden a estar correlacionados, con grupos al nivel del individuo.

En el experimento Misisipi los grupos son los distritos de la Reserva Federal. Hay sólo dos, lo que exige cautela. La autocorrelación tal vez no sea un problema en el experimento Misisipi, pero, de haberlo sido, entonces se habrían necesitado más datos antes de afirmar algo concluyente acerca de los efectos de la liquidez sobre la superviven-

cia de los bancos. Cuando se empiezan a definir grupos, la teoría formal que subyace a la inferencia estadística supone que se dispone de muchos grupos, en lugar de (o además de) muchas observaciones individuales dentro de cada grupo. En la práctica, «muchos» puede significar sólo unas cuantas docenas, como sucede en el caso de los estados de Estados Unidos. Puede que esto sea suficiente, pero un par o un puñado de grupos podrían no bastar.[14]

Los errores típicos robustos por grupos son adecuados en una gran variedad de contextos, no sólo para datos tabulados. En principio, los grupos resuelven cualquier tipo de problema de dependencia en los datos (aunque quizá los abultados errores típicos resultantes no sean de su agrado). Por ejemplo, es probable que los datos de las notas obtenidas por los estudiantes estén autocorrelacionados dentro de cada clase, si el alumnado de cada clase comparte el mismo profesor y tiene contextos familiares parecidos. Cuando se comunican las estimaciones de los efectos de determinadas actuaciones, como el efecto de los compañeros de clase en la ecuación (4.6) o los efectos de estudiar en universidades privadas del capítulo 2, los maestros calculan sus errores típicos robustos por grupos a nivel de clase, escuela o universidad.

[14] Véase un tratamiento más detallado de este punto en nuestro libro *Mostly Harmless Econometrics*, Princeton University Press, 2009. Andrew Jalil añadió grupos al experimento Misisipi en su análisis de cientos de condados a ambos lados de la frontera entre los distritos de la Reserva Federal. Véase «Monetary Intervention along the Atlanta Federal Reserve District Border», *Journal of Economic History*, vol. 74, número 1, marzo de 2014, páginas 259-273.

6

# El valor de la enseñanza

Habla la leyenda sobre un econometrista legendario cuyas habilidades econométricas eran dignas de leyenda.

***Maestros en acción***

Este capítulo completa nuestra exploración de los caminos que van de la causa al efecto con una investigación polifacética del efecto causal de la formación académica sobre los ingresos. Los cimientos de nuestro trabajo son las buenas preguntas, y el interrogante de si prolongar la escolaridad realmente incrementa los ingresos es ya un clásico. Es irónico, pero los maestros han tratado la cuestión educativa con todas las herramientas a su alcance excepto la asignación aleatoria. Las respuestas que han recopilado no son menos interesantes por el hecho de ser incompletas.

## 6.1 Formación académica, experiencia e ingresos

El veterano británico de la Segunda Guerra Mundial Bertie Gladwin abandonó la escuela secundaria a los 14 años de edad, y aun así encontró trabajo como ingeniero de radiocomunicaciones en los servicios de inteligencia británicos. Ya cumplidos los 60 regresó a las aulas y se graduó en psicología. Más tarde se graduó en microbiología,

estudios que culminó a la edad de 91 años. Desde entonces Bertie estuvo pensando si hacer el doctorado.[1]

Nunca es tarde para aprender cosas nuevas. Sin embargo, y a diferencia de Bertie Gladwin, la mayoría de los estudiantes completan sus estudios antes de asentar una carrera profesional. Los estudiantes universitarios pasan años enterrados entre libros y facturas de enseñanza, mientras que muchos de sus amigos de la escuela secundaria que no fueron a la universidad empiezan a trabajar y logran una cierta independencia económica. Los graduados universitarios esperan que el tiempo, el esfuerzo y el dinero dedicados, se vean recompensados al final del camino con un salario mayor. Las esperanzas y los sueños son una cosa, pero la vida depara muchas sorpresas. ¿Valen la pena los ingresos perdidos y los costes académicos asociados a la obtención de un título universitario? Esta es la pregunta del millón, y nuestro interés por ella es más que personal. Los contribuyentes subvencionan los estudios universitarios en todo el mundo, una política motivada en parte por la creencia de que la universidad es la clave del éxito económico.

Los economistas se suelen referir al efecto causal de los estudios sobre los ingresos como los *rendimientos de la formación académica*. Este término alude a la idea de que la escolaridad constituye una inversión en capital humano, con una contraprestación similar a la que brindan las inversiones financieras. Generaciones de maestros han estimado los rendimientos económicos de la formación. Sus esfuerzos ilustran cuatro de nuestras herramientas: regresión, DD, VI y RD.

El maestro de la econometría Jacob Mincer fue de los primeros en intentar cuantificar los rendimientos de la formación académica por medio de la regresión.[2] Mincer trabajó a partir de datos del censo de población de Estados Unidos y efectuó regresiones como

$$\text{en } Y_i = \alpha + \rho S_i + \beta_1 X_i + \beta_2 X_i^2 + e_i, \tag{6.1}$$

[1] Véase «"I'm Just a Late Bloomer": Britain's Oldest Student Graduates with a Degree in Military Intelligence Aged 91», *The Daily Mail*, 21 de mayo de 2012.

[2] El trabajo de Mincer aparece en su memorable libro *Schooling, Experience, and Earnings*, Columgia University Press y National Bureau of Economic Research, 1974.

donde $\ln Y_i$ es el logaritmo de los ingresos anuales del sujeto $i$, $S_i$ es su tiempo de escolaridad (medido en años), y $X_i$ los años de experiencia laboral. Mincer definía esta última variable como la edad menos los años de escolaridad menos 6, un cálculo que cuenta como trabajados todos los años tras la graduación. Los maestros llaman a la $X_i$ calculada de este modo la *experiencia potencial*. Es costumbre incorporar al modelo un control en forma de función cuadrática de la experiencia potencial, para tener en cuenta el hecho de que, aunque los ingresos crecen con la experiencia, lo hacen a un ritmo cada vez menor hasta acabar estabilizándose con la edad.

Las estimaciones de Mincer de la ecuación (6.1) para una muestra de unos 31.000 varones blancos urbanos en el censo de 1960 eran de este tipo:

$$\text{en } Y_i = \alpha + \underset{(0,002)}{0,070}\ S_i + e_i$$

$$\text{en } Y_i = \alpha + \underset{(0,002)}{0,107}\ S_i + 0,081\ X_i - 0,0012\ X_i^2 + e_i. \qquad (6.2)$$

Sin controles, $\rho = 0,07$. Esta estimación procede de un modelo logarítmico, así que $\rho = 0,07$ significa que los ingresos promedio crecen en torno a un 7% por cada año adicional de escolaridad (el apéndice del capítulo 2 trata los modelos con logaritmos en el primer miembro). Cuando se incluye la experiencia potencial como variable de control, los rendimientos estimados de la formación académica crecen hasta aproximadamente 0,11.

El modelo que tiene en cuenta la experiencia potencial controla por el hecho de que las personas con más escolaridad suelen tener menos años de experiencia laboral, puesto que suelen empezar a trabajar a tiempo completo más tarde (es decir, una vez acabados los estudios). Como $S_i$ y $X_i$ están correlacionados negativamente, la fórmula del SVO nos dice que si se omitiera la experiencia, que ejerce un efecto positivo sobre los ingresos, esto conduciría a una estimación de los rendimientos de la formación inferior a los que cabría esperar con el modelo largo, que incluye el control de la experiencia. Las estimaciones de Mincer indican que los varones blancos con un nivel de experiencia determinado disfrutan de un aumento de sus ingresos

en un 11% por cada año adicional de formación. Pero, a todas estas, falta comprobar si se trata de un efecto causal.[3]

### *Cantantes, espadachines y doctorados: el sesgo de aptitud*

La ecuación (6.1) compara entre sí a hombres con más y menos años de escolaridad, manteniendo fijos sus años de experiencia laboral. ¿Basta el control de la experiencia potencial para que *ceteris* sea *paribus*? Dicho de otro modo, a un nivel de experiencia dado, ¿los trabajadores con más formación son igual de diligentes y de capaces que los menos formados? ¿Tienen los mismos entornos familiares, capaces de darles un empujón en el mercado de trabajo? Unas respuestas afirmativas a estas preguntas serían difíciles de aceptar. Como otros maestros, nosotros estamos hiperformados. Y somos más listos, trabajamos más y tenemos entornos familiares más favorables que quienes no aguantaron tanto tiempo educándose, o eso nos decimos a nosotros mismos. Esos rasgos positivos que imaginamos que compartimos con otros trabajadores altamente formados se asocian también con unos mayores ingresos, lo que complica la interpretación causal de las estimaciones mediante regresión como las de la ecuación (6.2).

Por ello cabe esperar que mejoren las estimaciones de esta regresión simple si se añaden controles para otros atributos correlacionados con la escolaridad, variables a las que llamaremos $A_i$ (de «aptitud»). Si de momento nos olvidamos del término de la experiencia y nos centramos en otras fuentes de SVO, la regresión larga resultante se podría escribir como:

---

[3] La relación entre experiencia e ingresos descrita en estas estimaciones refleja un declive gradual del crecimiento de los ingresos con la edad. Para ver eso supongamos que incrementamos $X_i$ desde un valor $x$ hasta otro $x + 1$. El término $X_i$ crece en 1, mientras que $X_i^2$ crece en:

$$(x + 1)^2 - x^2 = 2x + 1.$$

El efecto neto de un incremento de un año en la experiencia es, pues:

$$(0{,}081 \times 1) - [0{,}0012 \cdot (2x + 1)] = 0{,}08 - 0{,}0024x.$$

Se estima, por tanto, que el primer año de experiencia incrementa los ingresos en casi un 8%, mientras que el décimo año de experiencia lo hace en tan sólo alrededor de un 5,6%. De hecho, el *perfil de experiencia*, como se denomina a esta relación, se torna completamente plano después de unos 30 años de experiencia.

$$\text{en } Y_i = \alpha^l + \rho^l S_i + \gamma A_i + e_i. \tag{6.3}$$

La fórmula del SVO nos dice que la pendiente de un modelo de regresión simple sin controles, $\rho^s$, se relaciona con la pendiente del modelo de regresión larga (6.3) mediante la fórmula

$$\rho^s \rho^l + \underbrace{\delta_{AS}\gamma}_{\text{sesgo de aptitud}},$$

donde $\delta_{AS}$ es la pendiente de una regresión simple de $A_i$ sobre $S_i$. Como siempre, simple ($\rho^s$) es igual a larga ($\rho^l$) más la regresión de la omitida (en la simple) sobre la incluida ($\delta_{AS}$) multiplicada por el efecto de la omitida en la larga ($\gamma$). En este contexto, la diferencia entre simple y larga se denomina *sesgo de aptitud*, porque la variable omitida es la aptitud.

¿En qué sentido juega el sesgo de aptitud? Hemos definido $A_i$ de manera que $\gamma$ en la regresión larga sea positivo (de otro modo, tendríamos que llamar a $A_i$ ineptitud). Sin duda, $\delta_{AS}$ es también positiva, lo que implica un sesgo de aptitud hacia arriba: por ello esperamos que el valor de la regresión simple $\rho_s$ supere al de la regresión más controlada, $\rho_l$. Al fin y al cabo, los estudiantes de la LSE y del MIT suelen tener grandes aptitudes, al menos en el sentido de lograr buenas notas en los exámenes y buenas calificaciones en secundaria. Por otra parte, hay gente que acorta su escolaridad por la mera razón de poderse dedicar cuanto antes a ganar dinero. Sir Mick Jagger abandonó su lucha por un título universitario en la LSE en 1963 para tocar en un grupo de música conocido como Rolling Stones. Jagger no encontró ninguna satisfacción en sus estudios y obviamente nunca se graduó en la facultad, pero ganó mucho como cantante de una banda de rock and roll. No es menos impresionante el caso del espadachín sueco Johan Harmenberg, que dejó el MIT en 1979 tras dos años de estudios y ganó una medalla de oro en esgrima en los Juegos Olímpicos de Moscú de 1980, en lugar de conseguir titularse en el MIT. Harmenberg se convirtió más tarde en un ejecutivo biotecnológico y un investigador de éxito. Estos ejemplos ilustran que las personas con grandes aptitudes (musicales, atléticas, empresariales o de otro tipo) pueden lograr el éxito económico sin contar con las ventajas de una formación académica. Esto en principio indica que $\delta_{AS}$ y, por tanto, el sesgo de aptitud tanto podría ser negativo como positivo.

### *La medida del hombre: el control de la aptitud*

He aquí un buen atajo para salvar la barricada del sesgo de aptitud: reunamos información acerca de $A_i$ y usémosla como control en las regresiones similares a la ecuación (6.3). En un esfuerzo por manejar el SVO en la estimación de los rendimientos de la formación, el maestro de la econometría Zvi Griliches usó el cociente intelectual (CI) como control de aptitud.[4] La estimación de Griliches de $\rho^s$ en un modelo con controles de la experiencia potencial, pero sin incluir el CI, asciende a 0,068. La estimación de Griliches del coeficiente de formación en su regresión corta cae bastante por debajo de la de Mincer, que rondaba el 11%, lo que podría deberse a las diferencias en cuanto a muestras y a variables dependientes (Griliches consideró efectos sobre ingresos por hora, y no por año). Pero lo importante es que la introducción de un control para el CI rebaja la estimación de Griliches hasta $\rho^l = 0{,}059$, una consecuencia del hecho de que el CI y la escolaridad estén fuertemente correlacionados de manera positiva, y que las personas con mayor CI ganen más (y, por tanto, el efecto de la aptitud omitida en el modelo largo es ciertamente positivo).

Por interesante que parezca, cuesta considerar definitivo el resultado de Griliches. El CI no capta el carisma de Mick Jagger ni la perseverancia de Johan Harmenberg, dimensiones de la aptitud que pocas veces se miden en las muestras estadísticas. El concepto relevante de aptitud equivaldría aquí a un *potencial de ingresos*, una idea que nos recuerda los resultados potenciales que se usan para describir los efectos causales a todo lo largo de este libro. El problema con los resultados potenciales es, como siempre, que nunca se pueden observar, que sólo podemos medir los resultados del camino tomado. Por ejemplo, sólo vemos el resultado potencial «altamente formado» en una muestra de graduados universitarios. No sabemos cómo les habría ido a estas personas si hubieran seguido a Johan y a Mick por el camino de irse de la universidad. Los intentos de resumir los ingresos potenciales con una variable única probablemente sean inadecuados. Además, por las razones que se explican en el apartado 6.2 y que se detallan mejor en el apéndice de este capítulo, cuando la formación

[4] Zvi Griliches, «Estimating the Returns to Schooling-Some Econometric Problems», *Econometrica*, vol. 45, número 1, enero de 1977, páginas 1-22.

académica se mide mal (y creemos que ocurre con frecuencia), las estimaciones en regresiones que incluyen controles de aptitud pueden resultar más pequeñas de lo que deberían ser.

### *Cuidado con un mal control*

Quizá la respuesta esté en incluir más controles. ¿Por qué no tener en cuenta la profesión, por ejemplo? Muchos conjuntos de datos que consignan ingresos clasifican también a los trabajadores de acuerdo con su dedicación, como ser jefe u obrero. Sin duda, la profesión es un predictor potente tanto del nivel académico como de los ingresos, y quizá incorpore los rasgos que diferencian a Mick y a Johan del ciudadano medio. Siguiendo entonces la lógica SVO tendríamos que controlar por profesiones, algo fácil de hacer si se incorporan variables binarias que marquen los tipos de oficios que se desempeñan.

Aunque la ocupación esté muy correlacionada tanto con la escolaridad como con el salario, las variables binarias de profesión son *controles malos* en aquellas regresiones que pretenden captar el efecto causal de la formación académica sobre los ingresos. El hecho de que el Maestro Joshway trabaje hoy día como profesor y no como auxiliar de enfermería (como hizo en tiempos) es en parte una recompensa por su extravagante formación académica. Sería un error eliminar este beneficio de nuestro cálculo y comparar únicamente entre profesores o entre auxiliares de enfermería cuando se trata de cuantificar el valor económico de la escolaridad. Incluso en un mundo en el que todos los profesores ganaran de manera uniforme un millón de dólares al año (esperemos el pronto advenimiento de ese mundo) y todos los auxiliares de enfermería de manera uniforme 10.000, un experimento que asignara niveles educativos al azar mostraría que la escolaridad incrementa los ingresos. El canal a través del cual crecen los salarios en este experimento imaginario sería el paso de la modesta ayudantía de enfermería al alto grado de profesor.

Hay, además, una segunda fuente de confusión, de carácter más sutil: los controles malos pueden ocasionar un sesgo de selección. Para ilustrarlo supongamos que nos interesan los efectos de haber obtenido un grado universitario, y que tal titulación se asigna de manera aleatoria. Los sujetos pueden ocupar uno de dos puestos

de trabajo, de cuello blanco o de cuello azul, y tener un grado universitario incrementa las probabilidades de tener el trabajo de cuello blanco. Como haberse graduado ejerce un efecto sobre el puesto de trabajo obtenido, la comparación de salarios dentro de cada estado de formación, condicionado al puesto de trabajo, ya no estaría bien equilibrada, incluso aunque los títulos universitarios se hayan asignado al azar y las comparaciones no condicionadas junten siempre manzanas con manzanas.

Este problemático fenómeno es lo que se suele denominar un *efecto de composición*. En virtud de la asignación aleatoria, quienes poseen y quienes no poseen un grado universitario son similares en todos los aspectos, al menos en promedio. Pero lo más importante es que tienen el mismo promedio $Y_{0i}$, es decir, el mismo potencial de ingresos promedio. Supongamos, sin embargo, que limitáramos la comparación a las personas cuyos oficios son de tipo cuello blanco. El grupo de control que no fue a la universidad constaría, en este caso, por completo de aquellos trabajadores excepcionales que lograron un trabajo de cuello blanco a pesar de no tener formación universitaria. El grupo de cuello blanco que se graduó en la universidad incluye también una porción de sujetos parecidos, que siempre habrían tenido trabajos de cuello blanco, más un grupo de trabajadores menos excepcionales que sólo habrían conseguido el trabajo de cuello blanco gracias a tener el título universitario, pero que no lo habrían logrado de otro modo.

Podemos ver las consecuencias de esta diferencia de composición si imaginamos tres grupos de trabajadores con el mismo número de sujetos. Los del primer grupo tienen el trabajo de cuello azul, hayan ido o no a la universidad (siempre azul, SA). Los del segundo grupo lucen cuello blanco con independencia de su formación (siempre blanco, SB). Los miembros del tercer grupo, azul blanco (AB), lucen cuello blanco solamente si tienen título universitario. Estos trabajos potenciales se describen en las dos columnas primeras de la tabla 6.1, que presenta los puestos de trabajo que obtendrían los sujetos de cada grupo en los escenarios alternativos de tener o no tener título universitario.

A pesar de que la asistencia a la universidad se asigne al azar, y aunque las comparaciones simples entre trabajadores con título universitario y sin él revelen efectos causales, las comparaciones dentro de

cada grupo de profesiones inducen a confusión. Supongamos, para aclarar el argumento, que el valor de haber ido a la universidad sea de 500 dólares semanales, por igual y para los tres grupos. Aunque los tres tipos de trabajadores disfruten de la misma ganancia por tener formación universitaria, es probable que sus ingresos potenciales difieran (es decir, sus valores de $Y_{0i}$). En concreto, supongamos que el grupo SB ganara 3.000 dólares a la semana sin título universitario, que el grupo SA ganara sólo 1.000 dólares semanales sin título, y que el grupo AB estuviera en medio, digamos en 2.000 dólares semanales sin título. Las columnas (3) y (4) de la tabla 6.1 resumen estas circunstancias.

**Tabla 6.1. Un mal control genera sesgos de selección**

| | **Profesión potencial** | | **Ingresos potenciales** | | **Ingresos promedio por profesión** | |
|---|---|---|---|---|---|---|
| | Sin título* | Con título* | Sin título* | Con título* | Sin título* | Con título* |
| Tipo de trabajador | (1) | (2) | (3) | (4) | (5) | (6) |
| Siempre azul (SA) | Azul | Azul | 1,000 | 1,500 | Azul 1,500 | Azul 1,500 |
| Azul blanco (AB) | Azul | Blanco | 2,000 | 2,500 | | Blanco 3,000 |
| Siempre blanco (SB) | Blanco | Blanco | 3,000 | 3,500 | Blanco 3,000 | |

* *Título universitario.*

Si la comparación entre quienes tienen título universitario y quienes no se restringe a los sujetos con trabajos de cuello blanco, entonces los ingresos promedio de los titulados universitarios vendrían dados por el promedio de los 3.500 dólares que ganan los SB con título y los 2.500 que ingresan los AB, mientras que el promedio de los no titulados son los 3.000 constantes que ganan los SB que no fueron a la universidad. Como el promedio de 3.500 y 2.500 también equivale a 3.000, la comparación por titulación universitaria condicionada a tener cuello blanco da cero, una estimación errónea de los rendimientos de la formación universitaria, que son 500 dólares para todo el mundo. La comparación de los ingresos según los estudios entre trabajadores de cuello azul vuelve a dar un valor equivocado y nulo.

Aunque la asignación aleatoria de estudios universitarios asegure proporciones iguales de manzanas y de peras (tipos o grupos) en las barricas de graduados y no graduados, condicionado a tener un trabajo de cuello blanco, un resultado determinado en parte por tener una educación universitaria, se distorsiona el equilibrio.

La moraleja del cuento del mal control es que el momento del tiempo resulta importante. Las variables medidas antes de que se determine el estado de tratamiento suelen ser buenos controles, porque el tratamiento no las puede alterar. Por el contrario, las variables de control que se miden tras el tratamiento pueden venir parcialmente determinadas por el propio tratamiento, en cuyo caso no son en absoluto controles, sino resultados. Introducir la profesión en un modelo de regresión que pretende calibrar el efecto causal de la formación académica es un caso muy claro. Los controles de aptitud, como las notas en los exámenes, podrían adolecer del mismo problema, sobre todo si las calificaciones proceden de exámenes que se realizan después de haber completado la mayor parte de la escolaridad. (Porque es probable que la escolaridad haga subir las calificaciones en esos exámenes.) Esta es una razón más para poner en duda las estrategias empíricas que confían en las notas de exámenes para suprimir el sesgo de aptitud en las estimaciones econométricas de los rendimientos de la formación académica.[5]

## 6.2 Los gemelos doblan la diversión

La ciudad de Twinsburg, en Ohio, cerca de Cleveland, se fundó con el nombre de Millsville a comienzos del siglo xix. Moses y Aaron Wilcox, prósperos empresarios de Millsville, eran gemelos idénticos a los que pocas personas podían distinguir. Sus éxitos animaron a Moses y Aaron a mostrarse generosos con Millsville, y esa actitud se vio recom-

[5] El lector atento se habrá percatado de que la experiencia potencial, que es una consecuencia a largo plazo de la escolaridad, también entra en la categoría de controles malos. En principio el sesgo podría eliminarse en este caso si se usara la edad y su cuadrado como instrumento para la experiencia potencial y su cuadrado. Como en los estudios a los que se hace referencia en el resto de este capítulo, también cabría simplemente sustituir el control de la experiencia por la edad, con lo que se pondría el objetivo sobre un efecto neto de la escolaridad que no se estima según las diferencias en experiencia potencial.

pensada, poco después, con el cambio del nombre de la ciudad, que pasó de ser Millsville («Villa de los Molinos») a llamarse Twinsbug («Ciudad de los Gemelos»). Desde 1975 Twinsburg honra su herencia cigótica con un festival de verano dedicado a aquellos gemelos. Las jornadas anuales de los gemelos de Twinsburg atraen no sólo a gemelos y mellizos que exhiben sus parecidos, sino también a investigadores en pos de comparaciones bien controladas.

Mellizos y gemelos tienen, de hecho, mucho en común: la mayoría se cría en la misma familia y al mismo tiempo, mientras que los gemelos, idénticos, comparten incluso los genes. Cabría afirmar, por tanto, que los gemelos comparten también las mismas capacidades genéticas. Cuando sucede que un gemelo tiene más formación académica que el otro, quizá se deba al tipo de fuerzas aleatorias tratadas en el capítulo 2. La idea de que un gemelo constituye un buen control para el otro motivó un par de estudios de los maestros Orley Ashenfelter, Alan Krueger y Cecilia Rouse.[6] La idea central de este trabajo, como en muchos otros que recurren a gemelos, consiste en que las aptitudes son comunes para cada par de hermanos, así que se pueden eliminar de la ecuación sustrayendo los datos de uno de los gemelos de los datos del otro y trabajando únicamente con las diferencias entre ambos.

La regresión larga que subyace a los análisis de los rendimientos de la formación utilizando datos de gemelos se pueden escribir como

$$\text{en } Y_{if} = \alpha^l + \rho^l S_{if} + \lambda A_{if} + e_{if}^l. \tag{6.4}$$

Aquí el subíndice $f$ significa «familia», mientras que el subíndice $i = 1, 2$ identifica a los gemelos, que pueden ser Amalia y Amelia, o Estanislao y Wenceslao. Si Wenceslao y Estanislao tienen las mismas aptitudes, entonces podemos simplificar la notación y escribir $A_{if} = A_f$. Esto a su vez implica que sus ingresos se pueden modelar como:

[6] Orley Ashenfelter y Alan B. Krueger, «Estimates of the Economic Returns to Schooling from a New Sample of Twins», *American Economic Review*, vol. 84, número 5, diciembre de 1994, páginas 1157-1173, y Orley Ashenfelter y Cecilia Rouse, «Income, Schooling, and Ability: Evidence from a New Sample of Identical Twins», *Quarterly Journal of Economics*, vol. 113, número 1, febrero de 1998, páginas 253-284.

$$\text{en } Y_{1,f} = \alpha^l + \rho^l S_{1,f} + \lambda A_f + e^l_{1f}$$

$$\text{en } Y_{1,f} = \alpha^l + \rho^l S_{2,f} + \lambda A_f + e^l_{2f}$$

Si se resta la ecuación de Estanislao a la de Wenceslao nos queda

$$\text{en } Y_{1,f} - \text{en } Y_{2,f} = \rho^l (S_{1,f} - S_{1,f}) + e^l_{2f} - e^l_{2f}, \tag{6.5}$$

una ecuación en la que desaparece la aptitud.[7] De aquí se deduce que cuando las aptitudes son constantes entre cada par de gemelos, una regresión simple de la diferencia de ingresos entre los gemelos sobre la diferencia en escolaridad recupera el coeficiente de la regresión larga, $\rho^l$.

Las estimaciones de la regresión construida sin efectuar la diferencia entre gemelos conduce a unos rendimientos de la formación en torno al 11%, bastante parecidos a los resultados de Mincer. Esto se aprecia en la primera columna de la tabla 6.2. El modelo que arroja los valores de la columna (1) incluye la edad, la edad al cuadrado, una variable binaria de sexo, y una variable binaria para raza blanca. Los gemelos blancos ganan menos dinero que los gemelos negros, un resultado infrecuente en el mundo de la comparación de ingresos por razas, aunque la diferencia en este caso no es significativamente distinta de cero.

La ecuación basada en las diferencias (6.5) da unos rendimientos de la formación alrededor del 6%, resultado que aparece en la columna (2) de la tabla 6.2. Esta cifra es sensiblemente inferior a la estimación de la regresión simple de la columna (1). Este declive podría reflejar cierto sesgo de aptitud en el modelo simple. Aun así, otra vez, podría haber fuerzas más sutiles en juego.

[7] Se pueden extraer estimaciones equivalentes a las de este modelo diferenciado si se añade una variable binaria para cada familia en un modelo en el que no se ejecuta la resta, y la muestra incluye a ambos gemelos. Las variables binarias familiares actúan como las de grupos selectivos en la ecuación (2.2) del capítulo 2, o como las de los estados de la ecuación (5.5) del apartado 5.2. Cuando hay sólo dos observaciones por familia, los modelos que se evalúan tras ejecutar la resta entre gemelos proporcionan una única observación por familia, pero generan estimaciones idénticas de los rendimientos de la formación idénticas a las que se obtienen cuando se marca con variables binarias a cada familia en una muestra conjunta que incluye a todos los gemelos.

### *Informes sobre gemelos de Twinsburg*

Los gemelos son similares en muchos sentidos, incluida (hete aquí) la escolaridad. De los 340 pares de gemelos entrevistados para los estudios de escolaridad de Twinsburg, alrededor de la mitad declaran niveles educativos idénticos. Las diferencias de escolaridad, $S_{1,f} - S_{2,f}$, varían mucho menos que los niveles de escolaridad $S_{if}$. Si la mayoría de los gemelos tienen realmente la misma formación, entonces es probable que un buen número de las diferencias de escolaridad que constan en los formularios se deban a errores de al menos uno de los gemelos al consignar los datos. Los formularios erróneos, llamados *errores de medida*, tienden a reducir las estimaciones de $\rho^I$ en la ecuación (6.5), un hecho que podría explicar la bajada de los rendimientos de la escolaridad estimados con el modelo diferenciado. Puede parecer poco importante que haya unos cuantos errores en los datos de escolaridad, pero las consecuencias de tales errores de medida pueden ser cruciales.

Para ver que estos errores son relevantes imaginemos que los gemelos de una misma familia tengan siempre la misma escolaridad. En este contexto, el único motivo por el que $S_{1,f} - S_{2,f}$ puede no ser nula es un error al comunicar los datos. Supongamos que esos formularios erróneos se producen al azar, debido a olvidos o a la falta de atención, en lugar de deberse a un efecto sistemático. El coeficiente de una regresión de las diferencias de ingresos sobre diferencias de escolaridad que son simplemente errores aleatorios debería ser cero, porque los errores al azar no guardan relación con los salarios. Pero en un caso intermedio, donde una parte de la variación observada en escolaridad, pero no toda, se deba a errores, el coeficiente de la ecuación (6.5) resultaría menor de lo que habría sido si la escolaridad se hubiera consignado bien. El sesgo inducido por este tipo de errores de medida en el regresor se llama *sesgo de atenuación* (*attenuation bias*). La fórmula matemática del sesgo de atenuación se deduce en el apéndice de este capítulo.

Los errores al comunicar la escolaridad atenúan las estimaciones de la regresión por niveles mostradas en la columna (1) de la tabla 6.2, pero menos que las estimaciones diferenciadas de la columna (2). Esta diferencia en el alcance del sesgo de atenuación se puede explicar con un escenario hipotético en el que todos los gemelos tuvieran

**Tabla 6.2. Rendimientos de la formación académica para los gemelos de Twinsburg**

| | Variable dependiente | | | |
|---|---|---|---|---|
| | Logaritmo de los ingresos | Diferencia en logaritmo de los ingresos | Logaritmo de los ingresos | Diferencia en logaritmo de los ingresos |
| Años de formación | .110<br>(.010) | | .116<br>(.011) | |
| Diferencia en años de formación | | .062<br>(.020) | | .108<br>(.034) |
| Edad | .104<br>(.012) | | .104<br>(.012) | |
| Cuadrado de la edad÷100 | –.106<br>(.015) | | –.106<br>(.015) | |
| Variable binaria para mujeres | –.318<br>(.040) | | –.316<br>(.040) | |
| Variable binaria para blancos | –.100<br>(.068) | | –.098<br>(.068) | |
| Instrumento para la formación vía informe de gemelos | No | No | Sí | Sí |
| Tamaño de la muestra | 680 | 340 | 680 | 340 |

*Notas:* Esta tabla revela las estimaciones sobre rendimientos de la formación para los gemelos de Twinsburg. La columna (1) presenta estimaciones mediante MCO del modelo de niveles. Las estimaciones MCO de los modelos con referencias cruzadas entre gemelos constan en la columna (2). La columna (3) da las estimaciones por MC2E de un modelo de regresión por niveles que usa el informe del hermano como instrumento para la escolaridad. La columna (4) refleja las estimaciones mediante MC2E usando la diferencia entre los informes de los hermanos como instrumento para la diferencia de escolaridad entre gemelos. Los errores típicos constan entre paréntesis.

la misma formación pero los niveles de escolaridad difirieran entre familias. Si los gemelos de una misma familia tienen siempre la misma escolaridad, entonces todas las diferencias de escolaridad dentro de las familias que aparezcan en los datos se deben a errores de medida. Pero, por el contrario, la mayoría de la variación entre familias que aparece en los datos se corresponde con diferencias reales en cuanto a formación. La variación real en formación está relacionada con los ingresos, un hecho que modera el sesgo de atenuación en las estimaciones del modelo por niveles, la ecuación (6.4). Esto refleja una cuestión general acerca de las consecuencias de las variables ex-

plicativas en modelos con regresores afectados por errores: los controles adicionales empeoran el sesgo de atenuación, un tema sobre el que se dan más detalles en el apéndice del capítulo.

Los errores de medida constituyen un problema importante en el análisis de Twinsburg, porque el error de medida por sí solo podría explicar el tipo de resultados que se observa en las columnas (1) y (2) de la tabla 6.2. Al pasar del estudio por niveles al modelo de regresión diferenciada se acentúa el sesgo de atenuación, y probablemente no poco. El descenso de los coeficientes de escolaridad al cambiar de columna podría tener poco que ver, por tanto, con el sesgo de aptitud. Por fortuna, los maestros experimentados Ashenfelter, Krueger y Rouse previeron el problema de atenuación. Pidieron a cada gemelo que informara no sólo de su propia escolaridad, sino también de la de su hermano. En consecuencia, el conjunto de datos de Twinsburg contiene dos medidas de la formación académica para cada sujeto, una del autoinforme y otra del informe escrito por el hermano. Los informes cruzados constituyen un contrapeso que reduce, y quizá incluso elimina, el sesgo de atenuación.

La herramienta clave en este caso, como en muchos otros problemas a los que nos hemos enfrentado, es la técnica VI. Amalia y Amelia cometen errores tanto al informar sobre el nivel de formación de la otra como al hacerlo del propio. Siempre que los errores cometidos por Amelia sobre la formación de su hermana carezcan de relación con los que su hermana comete en su autoinforme, y viceversa, el informe de Amalia sobre la escolaridad de Amelia puede usarse como un instrumento para el autoinforme de Amelia, y viceversa. El método VI elimina el sesgo de atenuación, tanto en la regresión por niveles como en las estimaciones del modelo diferenciado (aunque sigue siendo más probable que la regresión por niveles esté más afectada de sesgo de aptitud que la regresión diferenciada).

Como siempre, la estimación VI es el cociente entre las estimaciones de la forma reducida y las de la primera etapa. Al instrumentalizar la ecuación de niveles, la estimación de la forma reducida es el efecto del informe de Amalia acerca de la escolaridad de Amelia sobre los ingresos de esta última. La estimación correspondiente de la primera etapa es el efecto del informe de Amalia acerca de la escolaridad de Amelia sobre el autoinforme de esta última. Los resultados de la forma reducida y de la primera etapa siguen estando afectados de sesgo

de atenuación. Pero cuando dividimos uno entre el otro, estos sesgos se compensan y nos brindan una estimación VI no atenuada.

La técnica VI funciona de un modo parecido en el modelo de diferencias. El instrumento para las diferencias de escolaridad intrafamiliares es la diferencia entre los informes cruzados de los gemelos. Si los errores de medida en los autoinformes y en los informes cruzados no están correlacionados, entonces el método VI produce un valor en la regresión larga de los rendimientos de la formación, $\rho^{I}$, sin SVO y no atenuado, que es lo que querríamos obtener. La falta de correlación entre las equivocaciones en los informes de los hermanos es un supuesto fuerte, pero ofrece un punto de partida natural para cualquier exploración de los sesgos debidos a errores de medida.

La estimación VI de la ecuación de niveles aparece en la columna (3) de la tabla 6.2 (como siempre, el proceso VI se ejecuta por medio de MC2E, que funciona igual de bien con variables que no sean binarias). Al instrumentalizar la escolaridad de los autoinformes por medio de la escolaridad de los informes cruzados se incrementa la estimación de los rendimientos de la formación de 0,062 a 0,108. Este resultado consta en la columna (4) de la tabla 6.2 y apunta hacia unos errores de medida considerables en los datos de las diferencias. A la vez, la estimación VI con diferencias de 0,108 no queda muy por debajo de la estimación seccionada de 0,116, lo que sugiere que el problema que queremos resolver (el sesgo de aptitud en la estimación de los rendimientos de la formación) no es para tanto, al fin y al cabo.

## 6.3 Econometristas: por sus instrumentos los conocerás

### *Esa es la ley*

Los economistas creen que la gente toma decisiones importantes, como las referidas a la formación académica, comparando previsiones de costes con beneficios esperados. El coste de cursar estudios secundarios se determina en parte mediante las leyes de escolarización obligatoria que castigan a quienes dejan los estudios demasiado pronto. Como continuar con los estudios evita el castigo, las leyes de escolarización obligatoria hacen que la formación adicional parezca

más barata que su alternativa, que es dejar de estudiar. Esto genera una reacción causal en cadena que va desde las leyes de escolarización obligatoria hasta las decisiones de escolarización y de ingresos, lo que podría servir para revelar los rendimientos económicos de la formación académica. Los métodos econométricos basados en esta idea son los de los capítulos 3 y 5: variables instrumentales y diferencias en diferencias.

Como siempre, el proceso VI empieza con la primera etapa. Hace cien años había pocas leyes de escolarización obligatoria, mientras que hoy día la mayoría de los estados de Estados Unidos obligan a los jóvenes a permanecer escolarizados hasta al menos los 16 años de edad. Además hay muchos estados que prohíben que las personas en edad escolar trabajen, o que exigen que las autoridades educativas concedan en esos casos el permiso para trabajar. Si se supone que algunos estudiantes dejarían los estudios de no ser por estas leyes de escolarización obligatoria, se deriva que tales leyes incrementan la escolaridad promedio. Si fuera el caso que los cambios en las leyes de escolarización obligatoria carecen de relación con los ingresos potenciales de los residentes en cada estado (que vendrían determinados por circunstancias como el contexto familiar, la estructura industrial del estado, u otras diferencias políticas), entonces estas leyes constituyen instrumentos válidos para la escolaridad en ecuaciones como la (6.1).

Pero es probable que las leyes de escolarización obligatoria guarden relación con los ingresos potenciales. A comienzos del siglo xx, por ejemplo, los estados agrícolas del sur de Estados Unidos planteaban pocas exigencias de escolarización obligatoria, mientras que las leyes al respecto eran más estrictas en el norte, más industrializado. La comparación simple de los ingresos entre regiones de Estados Unidos suelen poner de manifiesto diferencias grandes, pero que en su mayor parte carecen de relación con el mayor rigor de los requisitos de escolarización en el norte. Las normas de escolarización obligatoria también se han vuelto más estrictas con el tiempo, pero de nuevo ocurre que las comparaciones simples conducen a error. Muchos rasgos de la economía estadounidense han cambiado en el transcurso del siglo xx, y las leyes de escolarización obligatoria no son sino una pequeña parte de esta historia económica en evolución permanente.

Una combinación creativa de DD y VI ofrece un posible atajo para evitar la barrera SVO en este contexto. Las exigencias de escolarización obligatoria se han extendido y se han endurecido del modo más espectacular en la segunda mitad del siglo xx. Los maestros Joshway y Daron Acemoglu recopilaron información anual para cada estado acerca de las leyes de escolarización obligatoria aplicables a quienes hubieran ido a la escuela en esos años.[8] Estas leyes incluyen disposiciones sobre el trabajo infantil, así como requisitos de asistencia obligatoria a las escuelas. Las leyes de trabajo infantil que exigen la superación de cierto nivel formativo antes de que los jóvenes puedan trabajar parecen haber incrementado la escolaridad más que la obligatoriedad de la asistencia. Una simplificación útil en este contexto utiliza las leyes en vigor en los estados de nacimiento de los censados en el momento en que éstos tenían 14 años de edad, para localizar así los estados y años en que se requerían siete, ocho, nueve o más años de escolaridad para autorizar la incorporación al mercado laboral. El conjunto resultante de variables instrumentales consiste en variables binarias para cada una de estas tres categorías, mientras que la categoría omitida corresponde a los estados en los que se requerían seis años de escolaridad o menos para poder trabajar.

Como los instrumentos de trabajo infantil varían con el estado y el año de nacimiento, se pueden usar para estimar una ecuación de primera etapa que controle posibles efectos temporales mediante la inclusión de variables binarias para el año de nacimiento, y que controle por las características de cada estado mediante variables binarias estatales. Al controlar el estado debería atenuarse el sesgo debido a diferencias regionales correlacionadas con las normas de escolarización obligatoria, mientras que al tener en cuenta efectos relacionados con el año de nacimiento se tendría que mitigar el sesgo derivado del hecho de que los ingresos difieren entre cohortes por muchos motivos distintos a las leyes de escolarización. La ecuación de la primera etapa resultante se parece al modelo de regresión DD del capítulo 5 (descrito en la ecuación [5.5]), empleado para estimar el efecto de los cambios estatales y anuales de la MLDA sobre

[8] Daron Acemoglu y Joshua D. Angrist, «How Large Are Human-Capital Externalities? Evidence from Compulsory-Schooling Laws», en Ben S. Bernanke y Kenneth Rogoff (editores), *NBER Macroeconomics Annual 2000*, volumen 15, MIT Press, 2001, páginas 9-59.

la mortalidad. Pero aquí las variables binarias de año de nacimiento sustituyen a las que allí marcaban años de calendario.

La ecuación de primera etapa sobre escolarización obligatoria de Acemoglu y Angrist se estimó con una extracción de varones en la cuarentena tomada de cada una de las muestras disponibles de los censos de Estados Unidos para cada década entre las de 1950 y 1990. Al acumular estos cinco censos se genera un único conjunto de datos en que los distintos censos contribuyen a diferentes cohortes. Por ejemplo, los hombres en la cuarentena observados en el censo de 1950 nacieron entre 1900 y 1909 y se someten a las leyes en vigor en las décadas de 1910 y 1920, mientras que los hombres en la cuarentena observados en el censo de 1960 nacieron entre 1910 y 1919 y para ellos rigen las leyes que estaban vigentes en las décadas de 1920 y 1930.

Las estimaciones de primera etapa que constan en la columna (1) de la tabla 6.3 parecen indicar que las leyes de trabajo infantil que requerían siete u ocho años de escolaridad antes de acceder al mercado laboral incrementaron la escolaridad (medida como el curso más alto completado) en unas dos décimas de año. Las leyes que requerían nueve o más años de formación académica antes de trabajar ejercieron un efecto el doble de grande. En la columna (3) de la misma tabla aparece un conjunto paralelo de estimaciones de la forma reducida. Estas resultan de modelos de regresión similares a los utilizados para construir las estimaciones de primera etapa de la columna (1), con el logaritmo de los ingresos semanales en lugar de los años de escolaridad como variable dependiente. Las leyes que exigen siete u ocho años de estudios para trabajar parecen incrementar los ingresos en un 1%, mientras que las que requieren nueve o más años lo hacen en casi un 5%, aunque sólo esta última estimación es significativa. La estimación MC2E generada por estos valores es 0,124 (con un error típico calculado de 0,036).

Un 12% de incremento salarial por cada año adicional de escolaridad es impresionante, y aún más si se tiene en cuenta que el incremento de la escolaridad en cuestión es involuntario. Las leyes más duras de escolarización obligatoria parecen incrementar la escolaridad, y esto a su vez proporciona ingresos más elevados para los varones sometidos a estas leyes (los cumplidores de la escolarización obligatoria, en este caso). Resulta especialmente interesante el hecho de que

la estimación MC2E de los rendimientos de la formación generada mediante instrumentos de escolarización obligatoria supere la estimación correspondiente por MCO, que es 0,075. Esta circunstancia contradice la idea de que pueda haber un sesgo positivo debido al sesgo de aptitud en la estimación MCO.

**Tabla 6.3. Rendimientos de la formación con el instrumento «trabajo infantil»**

| | Variable dependiente | | | |
|---|---|---|---|---|
| | Años de formación académica | | Logaritmo de los ingresos semanales | |
| | (1) | (2) | (3) | (4) |
| **A. Estimaciones de la primera etapa y de la forma reducida** | | | | |
| El trabajo infantil requiere 7 años | .166<br>(.067) | –.024<br>(.048) | .010<br>(.011) | -.013<br>.011 |
| El trabajo infantil requiere 8 años | –.191<br>(.062) | .024<br>(.051) | .013<br>(.010) | .005<br>.010 |
| El trabajo infantil requiere 9 o más años | .400<br>(.098) | .016<br>(.053) | .046<br>(.017) | .008<br>.014 |
| **B. Estimaciones de la segunda etapa** | | | | |
| Años de formación académica | | | .124<br>(.036) | .399<br>(.360) |
| Variable binaria de estado de nacimiento × tendencia lineal según año de nacimiento | No | Sí | No | Sí |

*Notas:* Esta tabla muestra las estimaciones mediante MC2E de los rendimientos de la formación académica usando como instrumentos tres variables binarias que marcan los años de escolaridad exigidos por las leyes para permitir el trabajo infantil. El apartado A refleja las estimaciones de la primera etapa y de la forma reducida, con controles para los efectos del año y el estado de nacimiento, así como variables binarias para el año del censo. Las columnas (2) y (4) muestran los resultados de añadir a la lista de controles tendencias lineales estatales específicas. El apartado B presenta las estimaciones por MC2E de los rendimientos de la formación generados por la primera etapa y la forma reducida del apartado A. El tamaño de la muestra es 722 343. Los errores típicos constan entre paréntesis.

Antes de proclamar misión cumplida, un maestro busca posibles amenazas a la validez de sus cálculos. La variación en escolaridad debida a las leyes de obligatoriedad genera una primera etapa y una forma reducida de estilo DD. Como se trata en el capítulo 5, la amenaza principal a la validez en este contexto está en la omisión de ten-

dencias estatales específicas. En concreto, debería preocuparnos el hecho de que los estados en los que las leyes de escolarización obligatoria se hicieron más estrictas experimentaron un crecimiento de los salarios entre cohortes muy grande, por razones sin relación con la formación. Quizá el crecimiento de los salarios y los cambios en las leyes educativas estén impulsados a la vez por alguna tercera variable como, por ejemplo, los cambios en la estructura industrial.

La sospecha de sesgo de variables omitidas en este contexto incluso aumenta si nos percatamos de que buena parte de la investigación sobre escolarización obligatoria se basan en una comparación entre los estados del norte y los del sur. Los estados del sur experimentaron un crecimiento económico enorme en el siglo xx, a la vez que proliferaba la legislación social en estos territorios. El incremento relativo de los ingresos en los estados del sur podría estar causado en parte por una normativa de escolarización más restrictiva. Pero también podría no ser así.

El capítulo 5 propone una comprobación sencilla de las tendencias estatales específicas por medio de una variable lineal para cada estado del modelo de interés. En este caso la dimensión temporal relevante sería el año de nacimiento, así que el modelo con tendencias estatales específicas incluye una variable lineal propia de cada estado de nacimiento (el modelo de regresión con tendencias según el año de nacimiento se parece a la ecuación (5.6)).

Las columnas (2) y (4) de la tabla 6.3 reúnen los resultados de este añadido. Las estimaciones de estas columnas brindan pocos indicios de que las leyes de escolarización obligatoria sean relevantes ni para la tasa de escolaridad ni para los ingresos. La estimación MC2E generada por las columnas (2) y (4) arroja un valor inverosímil por excesivo, 0,339, pero con un error típico casi igual de grande. Es triste decir al maestro Joshway que la tabla 6.3 refleja un diseño de investigación fallido.

### *Todo tiene su momento (de nacimiento)*

Maestro Oogway: El ayer es historia y el mañana un misterio, pero el hoy es un regalo. Por eso lo llamamos «presente».

*Kung Fu Panda*

Recibimos presentes en los cumpleaños, pero hay fechas de nacimiento mejores que otras. Un cumpleaños cercano a la Navidad puede reducir los obsequios si quienes te regalan intentan darte una sola cosa para cubrir a la vez ambas obligaciones. Por otra parte, muchos estadounidenses nacidos a finales de año reciben regalos sorpresa consistentes en una mayor formación académica o en unos ingresos más elevados.

El camino que conduce desde un nacimiento a final de año hasta una escolaridad o unos salarios mayores arranca en la guardería. En la mayoría de los estados se ingresa en la guardería en el año en que se cumplen los cinco de edad, con independencia de si se han cumplido ya o no en el momento de empezar el curso a comienzos de septiembre. Enrique, nacido el 1 de enero, iba ya camino de su sexto cumpleaños cuando entró en la escuela infantil. En contraste Diego, nacido el 1 de diciembre, ni siquiera había cumplido los cinco cuando empezó. A algunas personas estas diferencias en el inicio de la educación debidas a la fecha de nacimiento les han cambiado la vida.

El poder que tiene la edad a la que se empieza el colegio para cambiarnos la vida es una consecuencia no intencionada de las leyes estadounidenses de escolarización obligatoria. A mediados del siglo xx la mayoría de los estados de la unión permitían dejar los estudios (abandonar la escuela secundaria) sólo si ya se habían cumplido los 16 años (algunos estados exigían la escolarización hasta los 17 o los 18). La mayoría de leyes de escolarización obligatoria permiten dejar de estudiar si se ha alcanzado cierta edad límite aunque no se haya terminado el curso escolar. Enrique entró en el sistema educativo a la provecta edad de cinco años y ocho meses, y cumplió 16 en enero, 10 años después, cuando empezaba su décimo curso escolar. Diego ingresó en la escuela infantil con tan sólo cuatro años y nueve meses y alcanzó los 16 11 años más tarde, habiendo terminado el décimo curso y teniendo empezado ya el undécimo. Ambos estaban ansiosos por huir de la escuela tan pronto como se lo permitieran, y los dos lo hicieron en el acto, al cumplir los 16 años de edad. Pero Diego, que empezó a estudiar más joven, se vio forzado a completar un curso escolar más que Enrique, por un accidente de nacimiento.

No se puede elegir el cumpleaños. Es posible que incluso nuestros padres y madres tuvieran difícil fijarlo. En definitiva, el momento de nacimiento está afectado por una aleatoriedad considerable, que se

parece a una asignación aleatoria experimental. En virtud de la naturaleza parcialmente aleatoria de las fechas de nacimiento, es probable que personas como Enrique y Diego tengan contextos familiares y talentos similares, aunque alcancen logros académicos muy distintos. Esto suena a escenario prometedor para aplicar VI y, en efecto, lo es.

Los maestros Joshway y Alan Krueger emplearon las diferencias en escolaridad generadas por el trimestre de nacimiento (TDN) para construir estimaciones VI de los rendimientos económicos de la escolarización obligatoria.[9] Angrist y Krueger analizaron grandes muestras públicas de los censos estadounidenses de 1970 y 1980, unas muestras similares a las empleadas por Acemoglu y Angrist. Estos archivos censales contienen datos sobre el TDN de los encuestados, algo que no es frecuente en bancos de datos de acceso público.

La figura 6.1 presenta la primera etapa del estudio TDN para los encuestados en el censo de 1980. Esta gráfica muestra la escolaridad promedio, por año y TDN, para varones nacidos en la década de 1930. La mayoría de los varones de estas cohortes terminaron los estudios secundarios, por lo que el promedio de su curso más alto completado está entre 12 y 13 años. La figura 6.1 exhibe un perfil sorprendente con dientes de sierra: los varones nacidos en fechas tempranas del año tienden a tener, en promedio, menos escolaridad que los nacidos en fechas tardías. Los dientes de sierra tienen una amplitud en torno a 0,15. Esto podría no parecer mucho, pero concuerda con la historia de Enrique y Diego. Entre los varones nacidos en la década de 1930, alrededor del 20% abandonó la escuela en el décimo curso o antes. Nacer en los últimos trimestres impone alrededor de 0,75 cursos adicionales de escolaridad a este 20%. El cálculo $0{,}2 \times 0{,}75 = 0{,}15$ da cuenta de las oscilaciones de la figura 6.1.

Como siempre, la VI es el cociente de la forma reducida entre la primera etapa correspondiente. La figura 6.2 representa la forma reducida TDN. No nos sorprende el aspecto plano que muestran en esta gráfica los ingresos de un año a otro. Los salarios al principio crecen bastante con la edad, pero luego el perfil tiende a estabilizarse cuando se alcanza la cuarentena. Pero es importante observar que los dientes de sierra que traza la escolaridad cuando se representa frente

[9] Joshua D. Angrist y Alan B. Krueger, «Does Compulsory School Attendance Affect Schooling and Earnings?», *Quarterly Journal of Economics*, vol. 106, número 4, noviembre de 1991, páginas 979-1014.

al TDN tienen un paralelismo en los ingresos medios. Los varones que nacen en fechas tardías del año no sólo reciben más formación que los nacidos en fechas tempranas, sino que también tienen más ingresos. La lógica VI atribuye el patrón dentado de los ingresos promedio en función del TDN al patrón similar seguido por la escolaridad promedio.

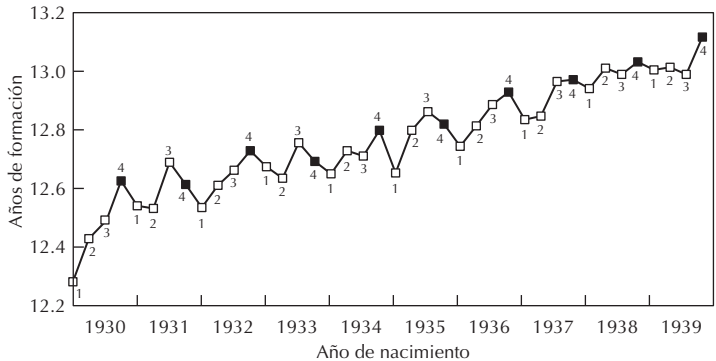



**Figura 6.1.** Primera etapa del trimestre de nacimiento.
*Notas:* Esta figura representa la escolaridad promedio en función del trimestre de nacimiento para varones nacidos entre 1930 y 1939 según datos del censo de 1980 de Estados Unidos. Los trimestres se etiquetan del 1 al 4, y los símbolos del cuarto trimestre aparecen sombreados.

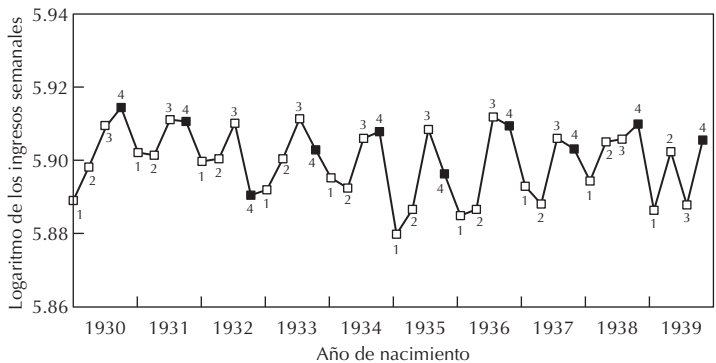



**Figura 6.2.** Forma reducida del trimestre de nacimiento.
*Notas:* Esta figura representa el promedio del logaritmo de los ingresos semanales en función del trimestre de nacimiento para varones nacidos entre 1930 y 1939 según datos del censo de 1980 de Estados Unidos. Los trimestres se etiquetan del 1 al 4, y los símbolos del cuarto trimestre aparecen sombreados.

Una estimación VI simple basada en el TDN compararía la escolaridad y los ingresos de varones nacidos en el cuarto trimestre con la escolaridad y los ingresos de varones nacidos en trimestres anteriores. La tabla 6.4 pone en orden los ingredientes de esta receta VI sobre la misma muestra de la figura 6.1. Los varones nacidos en el cuarto trimestre ganan un poco más que los nacidos antes, y la diferencia asciende a alrededor del 0,7%. Los nacidos en el cuarto trimestre logran también un mayor grado formativo promedio, con una diferencia en el entorno de 0,09 años. Si se divide la primera diferencia entre la segunda, tenemos:

$$\textit{Efecto de la escolaridad sobre los ingresos} = \frac{[\textit{Efecto del TDN sobre los ingresos}]}{[\textit{Efecto del TDN sobre la escolaridad}]}$$

$$= \frac{0{,}0068}{0{,}0092} = 0{,}074.$$

**Tabla 6.4. Receta VI para una estimación de los rendimientos de la formación académica con un solo instrumento «trimestre de nacimiento»**

| | Nacidos en trimestres 1-3 | Nacidos en trimestre 4 | Diferencia |
|---|---|---|---|
| Logaritmo de los ingresos semanales | 5.8983 | 5.9051 | .0068<br>(.0027) |
| Años de formación | 12.74736 | 12.8394 | .0921<br>(.0132) |
| Estimación VI de los rendimientos de la formación académica | | | .074<br>(.028) |

*Notas:* El tamaño de la muestra es 329.509. Se dan entre paréntesis los errores típicos.

A modo de comparación, la regresión simple del logaritmo de los ingresos semanales sobre la escolaridad llega a un resultado sensiblemente similar, 0,071. Estas estimaciones sencillas con MCO y VI se repiten en las dos primeras columnas de la tabla 6.5. Las columnas que ofrecen estimaciones VI están etiquetadas como «MC2E» porque, como siempre, la técnica VI se aplica de este modo.

**Tabla 6.5. Rendimientos de la formación académica con instrumentos «trimestre de nacimiento» alternativos**

| | MCO | MC2E | MCO | MC2E | MC2E |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Años de formación | .071<br>(.0004) | .074<br>(.028) | .071<br>(.0004) | .075<br>(.028) | .105<br>(.020) |
| Estimador estadístico *F* de la primera etapa | | 48 | | 47 | 33 |
| Instrumentos | Ninguno | Trimestre | Ninguno | Trimestre | Variables binarias para tres trimestres |
| Controles de año de nacimiento | No | No | Sí | Sí | Sí |

*Notas:* Esta tabla refleja las estimaciones mediante MCO y MC2E de los rendimientos de la formación académica utilizando el trimestre de nacimiento para construir instrumentos. Las estimaciones de las columnas (3) a (5) proceden de modelos que incluyen controles para el año de nacimiento. Las columnas (1) y (3) presentan las estimaciones MCO. En las columnas (2), (4) y (5) constan las estimaciones MC2E que usan los instrumentos indicados en la fila tercera de la tabla. En la segunda fila se da la prueba *F* sobre el significado conjunto de los instrumentos en las correspondientes regresiones de primera etapa. El tamaño de la muestra es 329.509. Se dan entre paréntesis los errores típicos.

Como en el caso de la estimación VI de los efectos del tamaño familiar tratados en el capítulo 3, podemos usar MC2E para añadir variables explicativas e instrumentos adicionales al procedimiento TDN mediante VI. Las estimaciones mediante MCO y MC2E con modelos que incluyen variables binarias para el año de nacimiento (un control de edad para nuestra sección de 1980) constan en las columnas (3) y (4) de la tabla 6.5. Estos resultados son casi indistinguibles de los que aparecen en las columnas (1) y (2). Sin embargo, al añadir a la lista de instrumentos unas variables binarias para marcar los trimestres primero y segundo se logra un incremento notable de la precisión. La estimación con tres instrumentos que aparece en la columna (5) es mayor que la que resulta con sólo uno y que consta en las columnas (2) y (4), con un error típico que cae desde 0,028 hasta 0,020.

¿Qué es lo que hace falta para que las estimaciones MC2E que usan instrumentos TDN capten el efecto causal de la formación aca-

démica sobre los ingresos? En primer lugar, los instrumentos deben predecir el regresor de interés (en este caso, la escolaridad). En segundo lugar, los instrumentos tienen que ser tan buenos como si se asignaran al azar, en el sentido de ser independientes de las variables omitidas (en este caso, variables como contexto familiar y aptitudes). Y por último, el TDN tiene que afectar a los ingresos sólo a través del canal elegido como variable instrumentalizada (en este caso, la escolaridad). No debe haber otros canales. Vale la pena preguntarse de qué manera los instrumentos TDN cumplen estos requisitos de primera etapa: independencia y restricción de exclusión.

El TDN ¿es independiente de las características maternales? Por supuesto, los cumpleaños no se asignan rigurosamente al azar. Los investigadores han documentado hace tiempo el carácter estacional de los patrones de nacimientos en función del contexto socioeconómico de las madres. Un estudio reciente de Kasey Buckles y Daniel Hungerman explora a fondo estos patrones.[10] Buckles y Hungerman descubren que la escolaridad de la madre (una buena medida del contexto familiar) alcanza un máximo en el caso de madres que dan a luz en el segundo trimestre. Esto sugiere que el contexto familiar no puede ser la causa del patrón estacional en educación e ingresos que se aprecia en las figuras 6.1 y 6.2, porque ambos exhiben máximos en los trimestres tercero y cuarto. De hecho, la escolaridad materna promedio en función del TDN presenta una ligera correlación negativa con la escolaridad promedio de la descendencia en función del TDN. Por tanto, no es sorprendente que, si se controlan las características maternas promedio, se incrementen de manera moderada las estimaciones VI de los rendimientos de la formación con instrumentos TDN. El cambio estacional de los nacimientos en función del contexto familiar no es nulo, pero no sigue un patrón capaz de alterar de manera sustancial las estimaciones por MC2E basadas en TDN.

[10] Kasey Buckles y Daniel M. Hungerman, «Season of Birth and Later Outcomes: Old Questions, New Answers», NBER Working Paper 14573, National Bureau of Economic Research, diciembre de 2008. Véase también John Bound, David A. Jaeger y Regina M. Baker, que fueron los primeros en llamar a atención sobre la posibilidad de que las estimaciones con instrumentos TDN no admitieran una interpretación causal en «Problems with Instrumental Variables Estimation When the Correlation between the Instruments and the Endogeneous Explanatory Variable Is Weak», *Journal of the American Statistical Association*, vol. 90, número 430, junio de 1995, páginas 443-450.

Finalmente, ¿qué hay de la exclusión? La primera etapa TDN viene generada por el hecho de que los estudiantes nacidos en fechas tardías del año acceden más jóvenes a la escuela que los que nacieron en fechas más tempranas, y por lo tanto completan más escolaridad antes de que se les permita dejar los estudios. Pero ¿qué sucedería si la edad de inicio de los estudios tuviera importancia por sí misma? La argumentación habitual al respecto afirma que el alumnado más joven de primer curso está en desventaja, mientras que quienes son algo mayores que sus compañeros de clase tienden a rendir mejor. Y de nuevo aquí los indicios circunstanciales a favor de los instrumentos TDN son reconfortantes. La clave de la hipótesis que relaciona el TDN con la escolarización obligatoria consiste en que quienes entran más jóvenes en la escuela acaban ganando más, y esto es lo que revelan los datos.[11]

Las estrategias empíricas nunca son perfectas. Los clavos endebles se doblan, pero la casa de la econometría no tiene por qué hundirse. No podemos demostrar que una estrategia VI concreta satisfaga los requisitos de una interpretación causal. La posición del económetra tiene que ser necesariamente defensiva. Pero, tal como hemos visto, es posible sondear y verificar los requisitos fundamentales de varias maneras, y eso es lo que debe hacerse. Los maestros siempre comprueban sus propios trabajos y supuestos, a la vez que evalúan con cuidado los resultados obtemidos por otros.

Yendo a lo sustancial, las estimaciones de los rendimientos económicos de la formación académica con técnicas VI que usan instrumentos TDN resultan similares o mayores que las estimaciones correspondientes por medio de MCO. Un error de medida modesto en la variable de escolaridad podría explicar la diferencia entre las estimaciones con MC2E y MCO, al estilo de lo que ocurría en los datos de gemelos. Estos resultados sugieren que el sesgo a la baja debido a errores en la medida de la escolaridad puede tener la misma importancia, o más, que cualquier sesgo de aptitud que nos llevara a sobreestimar el valor económico de la formación académica. La mejora de ingresos generada por un curso escolar

[11] Véanse más detalles acerca de este punto en Joshua D. Angrist y Alan B. Krueger, «The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables with Moments from Two Samples», *Journal of the American Statistical Association*, vol. 87, número 418, junio de 1992, páginas 328-336.

adicional completo parece situarse entre el 7 y el 10%. A Bertie Gladwin podría haberle ido todavía mejor si hubiera completado su escolaridad antes.

### 6.4 Flamantes badanas en el estado de la estrella solitaria

La escolaridad significa muchas cosas, y cada experiencia educativa es distinta. Pero los economistas observan las diversas experiencias educativas y las contemplan como creadoras de capital humano: una inversión cara en capacidades de la que se espera un rendimiento. Hay estudiantes, como Bertie Gladwin, que disfrutan de la formación en sí misma y que no muestran interés por su rendimiento económico. Pero probablemente haya muchas más personas que consideen sus estudios como estresantes, agotadores y caros. Aparte del coste de la matrícula, el tiempo invertido en los estudios se podría haber dedicado a trabajar. Muchos estudiantes universitarios gastan relativamente poco en los estudios en sí, pero todo estudiante a tiempo completo paga un coste de oportunidad. Esta noción de que una gran parte de los costes de recibir una formación consisten en los ingresos no percibidos, nos lleva a esperar que cada año de escolaridad adicional genere aproximadamente el mismo rendimiento económico, sea el décimo, el duodécimo o el vigésimo año pasado entre libros. Una visión simple de la escolarización desde el punto de vista del capital humano incorpora esta idea.

Por supuesto, las personas sin nociones de economía probablemente no se plantean la formación académica de esta manera. La mayoría mide sus logros educativos en términos de títulos obtenidos, y no de años invertidos. Pocos solicitantes de empleo se describen a sí mismos como alguien que «ha completado 17 años de escolaridad». Es más habitual proporcionar la lista de centros educativos en los que se ha logrado un título, y las fechas en las que se obtuvieron. Sin embargo, para un economista los títulos educativos no son más que trozos de papel con escaso o nulo valor real. Viene al caso el ejemplo del maestro Stevefu: aunque pasó muchos años en la facultad, en la Universidad Susquehanna del centro de Pensilvania (entre otras instituciones de excelencia), todavía no posee el título de grado universitario. Este menosprecio del valor de los títulos se refleja en la expre-

sión con que los economistas se refieren a la supuesta importancia de los títulos académicos, la cual llaman «efectos badana», en alusión al material en que se inscribían originalmente, piel de cordero.

La búsqueda de efectos badana condujo a los maestros Damon Clark y Paco Martorell a un ingenioso diseño de RD difusa.[12] Explotaron el hecho de que en Tejas, como en muchos otros estados, para obtener el diploma de secundaria hay que superar un examen final, aparte de superar durante el curso las competencias exigidas por el estado. Los estudiantes realizan este examen por primera vez en los cursos décimo o undécimo, y se planifican nuevas convocatorias para quienes suspenden. Al final del duodécimo curso se oferta un examen de última convocatoria para quienes suspendieron todos los anteriores. En realidad, esta no es la última oportunidad para que un veterano de secundaria de Tejas obtenga su diploma, porque puede volver a intentarlo más tarde. Aun así, para muchas personas este examen de última convocatoria resulta decisivo.

La naturaleza decisiva de esta última convocatoria para muchos veteranos de secundaria de Tejas se refleja en la figura 6.3, que representa la probabilidad de obtener el diploma en función de la nota en este examen, centrada alrededor de la nota de corte para el aprobado. La gráfica presenta promedios condicionados para cada calificación, así como los valores resultantes de ajustar un polinomio de cuarto grado estimado por separado a un lado y a otro de la nota de corte. Se observa que hay una tasa de obtención del diploma en torno a 0,5 para los estudiantes que quedan por debajo del umbral. Sin embargo, para quienes superan el umbral, la tasa de graduaciones da un salto que supera el 90%. Este cambio es discontinuo y no deja margen a la ambigüedad: la figura 6.3 documenta una primera etapa de RD difusa de alrededor de 0,5 para los efectos de aprobar el examen sobre la probabilidad de obtener el diploma.

Muchas personas que obtienen el diploma acuden a la universidad, en cuyo caso sus ingresos se mantienen bajos hasta que completan esta fase formativa adicional. Por eso es importante otear lo bastante lejos en el camino para buscar la aparición de algún efecto badana en los ingresos. Clark y Martorell usaron datos del sistema de

[12] Damon Clark y Paco Martorell, «The Signaling Value of a High School Diploma», *Journal of Political Economy*, vol. 122, número 2, abril de 2014, páginas 282-318.

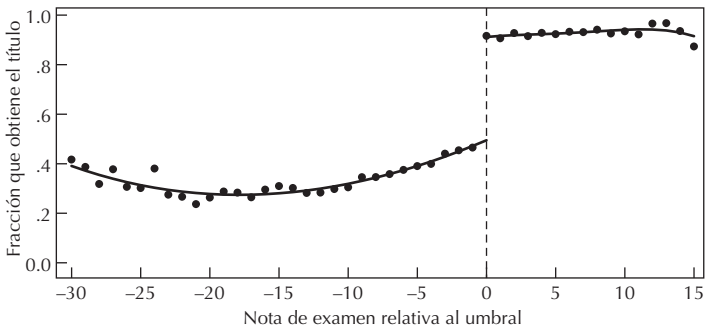



**Figura 6.3.** Notas en el examen de última convocatoria y titulaciones en Tejas.

*Notas:* La notas en el examen de última convocatoria están normalizadas respecto de los umbrales de aprobado. Los puntos muestran la fracción de sujetos que reciben el diploma para cada nota en el examen. Las líneas continuas representan los valores ajustados mediante polinomios de cuarto grado, estimados por separado a cada lado del umbral (marcado por la línea vertical discontinua).

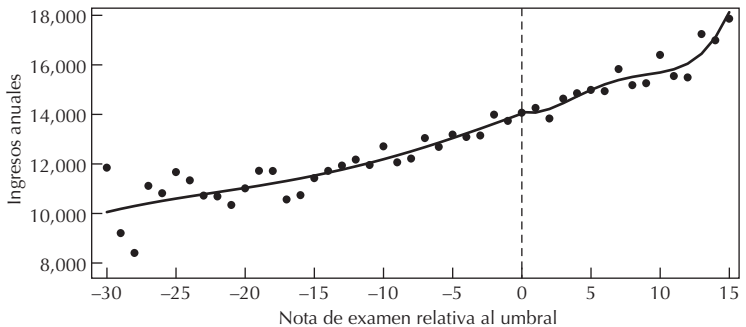



**Figura 6.4.** Efectos de la nota en el examen de última convocatoria sobre los ingresos.

*Notas:* Las notas en el examen de última convocatoria están normalizadas respecto de los umbrales de aprobado. Los puntos muestran los ingresos promedio para cada nota en el examen, incluyendo ceros para quienes no trabajan. Las líneas continuas representan los valores ajustados mediante polinomios de cuarto grado, estimados por separado a cada lado del umbral (marcado por la línea vertical discontinua).

seguros de desempleo de Tejas, que contiene información longitudinal sobre los ingresos de la mayoría de trabajadores del estado, para seguir la evolución a lo largo de 11 años de los ingresos de quienes hicieron el examen en última convocatoria.

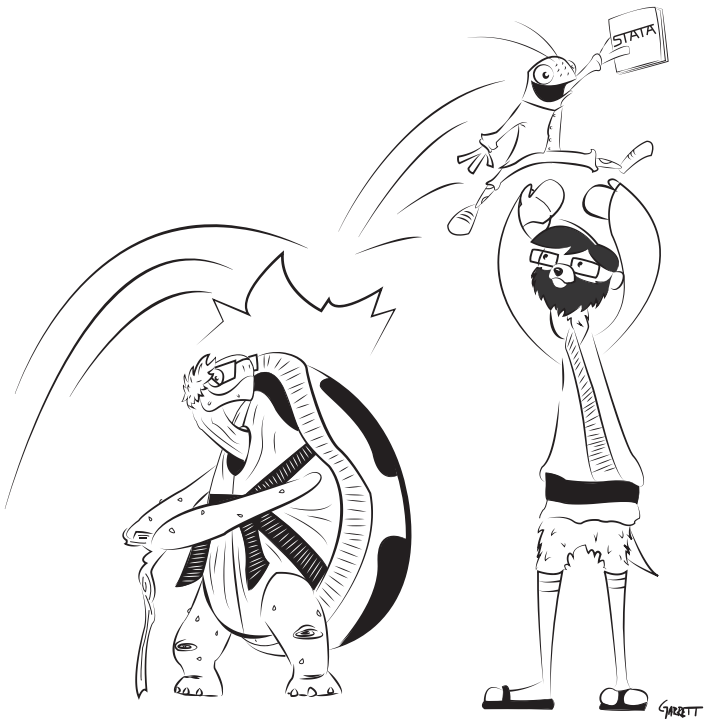
Los datos de ingresos de los estudiantes que hicieron el examen en última convocatoria a lo largo de entre 7 y 11 años no muestran ningún indicio de efectos badana. Esto se aprecia en la figura 6.4, que muestra los ingresos anuales promedio en función de las notas del examen, en un formato paralelo al de la figura 6.3 (los ingresos aquí constan en dólares, y no en su logaritmo, y los promedios incluyen ceros para las personas que no trabajan). La figura 6.4 constituye una representación gráfica de la forma reducida en un diseño de RD difusa que usa una variable binaria que indica si se aprobó el examen final en última convocatoria como instrumento para el efecto de la obtención del diploma sobre los ingresos. Como siempre, cuando la forma reducida es cero (en este caso no se aprecia un escalón en la figura 6.4) sabemos que la estimación MC2E resultante también será cero.

Las estimaciones MC2E generadas al dividir las discontinuidades apreciadas en la primera etapa y en la forma reducida de las figuras 6.3 y 6.4 muestran un efecto del diploma de 52 dólares (con un error típico en torno a 630 dólares). Esto asciende a menos de medio punto porcentual de los ingresos promedio, que rondan los 13.000 dólares. Se trata, de hecho, de efectos pequeños, lo que va en contra de la hipótesis badana. Por otra parte, los intervalos de confianza asociados incluyen también efectos sobre los ingresos que ascienden casi al 10%.

Unos errores típicos grandes nos dejan ante la posibilidad de que haya algunos efectos badana, así que la búsqueda de indicios acerca de esta cuestión sin duda continuará. Los maestros saben que la búsqueda de la verdad econométrica jamás cesa, y que lo que es bueno hoy será mejorado mañana. Nuestros alumnos nos lo enseñan.

Maestro Stevefu: Es hora de partir, Pequeño Saltamontes. Habrás de proseguir tu viaje en solitario. Recuerda que cuando se sigue el sendero de la econometría, cualquier cosa es posible.

Maestro Joshway: Cualquier cosa es *posible*, Pequeño Saltamontes. Pero aun así, mide siempre la evidencia.




## Apéndice: Sesgo debido a errores de medida

Soñamos con efectuar la regresión:

$$Y_i = \alpha + \beta S_i^* + e_i, \tag{6.6}$$

pero los datos acerca de $S_i^*$, el regresor de nuestros sueños, no están disponibles. Tan sólo contamos con una versión mal medida, $S_i$. Escribamos la relación entre ambos regresores de este modo:

$$S_i = S_i^* + m_i, \tag{6.7}$$

donde $m_i$ es el error de medida inherente a $S_i$. Para simplificar, supongamos que el error promedio es nulo y que no está correlacionado ni con $S_i^*$ ni con los residuos $e_i$. Entonces tenemos:

$$E[m_i] = 0$$
$$C(S_i^*, m_i) = C(e_i, m_i) = 0$$

Estos supuestos describen un *error de medida clásico* (errores de medida de estilos más coloridos podrían hacer bailar aún más los coeficientes de regresión).

El coeficiente de regresión que perseguimos, $\beta$ en la ecuación (6.6), viene dado por:

$$\beta = \frac{C(Y_i, S_i^*)}{V(S_i^*)}.$$

Si se emplea el regresor afectado de errores, $S_i$, en lugar de $S_i^*$, tenemos:

$$\beta_b = \frac{C(Y_i, S_i)}{V(S_i)}, \tag{6.8}$$

donde $\beta_b$ lleva el subíndice $b$ para recordarnos que está sesgado.

Que $\beta_b$ es una versión sesgada del coeficiente que buscamos se deduce empleando las ecuaciones (6.6) y (6.7) para sustituir $Y_i$ y $S_i$ en el numerador de la ecuación (6.8):

$$\begin{aligned}\beta_b &= \frac{C(Y_i, S_i)}{V(S_i)} \\ &= \frac{C(\alpha + \beta S_i^* + e_i, S_i^* + m_i)}{V(S_i)} \\ &= \frac{C(\alpha + \beta S_i^* + e_i, S_i^*)}{V(S_i)} = \frac{V(S_i^*)}{V(S_i)}.\end{aligned}$$

En la penúltima igualdad se recurre al supuesto de que el error de medida, $m_i$, no está correlacionado con $S_i^*$ ni con $e_i$, mientras que la última igualdad se basa en que $S_i^*$ no está correlacionada con una constante ni con $e_i$, dado que este último es el residuo de una regresión sobre $S_i^*$. También hemos utilizado el hecho de que la covarianza de $S_i^*$ consigo misma es su varianza (en el apéndice del capítulo 2 se ofrece una explicación de estas propiedades, y otras relacionadas, de la varianza y la covarianza).

Admitimos la falta de correlación entre $m_i$ y $S_i^*$. Como la varianza de una suma de variables no correlacionadas es la suma de sus varianzas, esto implica que:

$$V(S_i) = V(S_i^*) + V(m_i),$$

lo que significa que podemos escribir

$$\beta_b = r\beta, \tag{6.9}$$

donde

$$r = \frac{V(S_i)}{V(S_i)} = \frac{V(S_i^*)}{V(S_i^*) + V(m_i)}$$

es un número que toma valores entre cero y uno.

La fracción $r$ describe la proporción de la variación de $S_i$ que no está relacionada con los errores, y se denomina la *fiabilidad* de $S_i$. La fiabilidad determina hasta qué punto el error de medida atenúa $\beta_b$. El sesgo de atenuación en $\beta_b$ es:

$$\beta_b - \beta = -(1 - r)\beta,$$

de manera que $\beta_b$ es menor que $\beta$ (si es positivo), a menos que $r = 1$, en cuyo caso no hay errores de medida después de todo.

### *Añadir variables explicativas*

En el apartado 6.1 indicamos que al añadir variables explicativas a un modelo con regresores afectados por errores de medida se tiende a acentuar el sesgo de atenuación. El experimento de Twinsburg relatado en el apartado 6.2 constituye un caso especial de esto, donde las variables explicativas son variables binarias que indican las familias en las muestras de gemelos. Para ver de qué manera las variables explicativas acentúan el sesgo de atenuación, supongamos que la regresión de interés es:

$$Y_i = \alpha + \beta S_i^* + \gamma X_i + e_i, \tag{6.10}$$

donde $X_i$ es una variable de control, quizá el CI u otra nota de examen. Sabemos por la anatomía de la regresión que el coeficiente de $S_i^*$ en este modelo viene dado por

$$\beta = \frac{C(Y_i, \widetilde{S}_i^*)}{V(\widetilde{S}_i^*)},$$

donde $\widetilde{S}_i^*$ es el residuo de una regresión de $S_i^*$ sobre $X_i$. Del mismo modo, si se sustituye $S_i^*$ por $S_i$, el coeficiente de $S_i$ se convierte en

$$\beta_b = \frac{C(Y_i, \widetilde{S}_i)}{V(\widetilde{S}_i)},$$

donde $\widetilde{S}_i$ es el residuo de una regresión de $S_i$ sobre $X_i$.

Añadamos el supuesto (clásico) de que el error de medida, $m_i$, no está correlacionado con la variable $X_i$. Entonces el coeficiente de una regresión de la variable afectada de error $S_i$ sobre $X_i$ es el mismo que el coeficiente de una regresión de $S_i^*$ sobre $X_i$ (usando las propiedades de la covarianza y la definición de coeficiente de regresión), lo que a su vez implica que:

$$\widetilde{S}_i = \widetilde{S}_i^* + m_i,$$

donde $m_i$ y $\widetilde{S}_i^*$ no están correlacionadas. Por lo tanto, tenemos que:

$$V(\widetilde{S}_i) = V(\widetilde{S}_i) + V(m_i),$$

Si se aplica la lógica empleada para deducir la ecuación (6.9), obtenemos:

$$\begin{aligned} \beta_b &= \frac{C(Y_i, \widetilde{S}_i)}{V(\widetilde{S}_i)} \\ &= \frac{V(\widetilde{S}_i)}{V(\widetilde{S}_i^*) + V(m_i)}\beta = \overline{r}\beta, \end{aligned} \tag{6.11}$$

donde

$$\overline{r} = \frac{V(\widetilde{S}_i^*)}{V(\widetilde{S}_i) + V(m_i)}.$$

Como $r$, esta cantidad adopta valores entre cero y uno.

¿Qué hay de nuevo aquí? La varianza de $\widetilde{S}_i^*$ a la fuerza está mermada frente a la de $S_i^*$, porque la varianza de $\widetilde{S}_i^*$ es la varianza de un residuo de un modelo de regresión en el que $S_i^*$ es la variable dependiente. Como $V(\{S \text{ tilde}\}_i^*) < V(S_i^*)$, tenemos también que

$$\overline{r} = \frac{V(\widetilde{S}_i^*)}{V(\widetilde{S}_i^*) + V(m_i)} < \frac{V(S_i^*)}{V(S_i^*) + V(m_i)} = r.$$

Esto explica por qué al añadir variables explicativas a un modelo con errores en la medida de la escolaridad se agrava el sesgo de atenuación en la estimación de los rendimientos de la formación académica. La intuición llevaría a interpretar esta agravación como una consecuencia de que las variables explicativas están correlacionadas con la escolaridad bien medida, pero carecen de relación con los errores. La operación vía anatomía de la regresión que elimina la influencia de las variables explicativas reduce, por tanto, el contenido informativo de un regresor afectado de errores, pero deja inalterada la componente de ruido, los errores (ponga a prueba su comprensión del argumento formal derivando la ecuación [6.11]). Este argumento se traslada a la operación de resta que se usa para extraer la aptitud de la ecuación (6.4): la diferencia entre gemelos retira parte de la señal de escolaridad, pero deja inalterada la varianza debida a ruido.

### *VI nos despeja el camino*

Sin variables explicativas, la fórmula VI para el coeficiente $S_i$ en una regresión simple es:

$$\beta_{VI} = \frac{C(Y_i, Z_i)}{C(S_i, Z_i)}, \tag{6.12}$$

donde $Z_i$ es el instrumento. En el apartado 6.2, por ejemplo, usamos los informes cruzados entre hermanos como instrumento para la escolaridad, que podría constar erróneamente en el autoinforme. Si el instrumento no está correlacionado ni con el error de medida ni con los residuos, $e_i$, en ecuaciones como la (6.6), el método VI elimina el sesgo debido a errores de medida en $S_i$.

Para ver cómo trabaja el sistema VI en este contexto usemos las ecuaciones (6.6) y (6.7) para sustituir $Y_i$ y $S_i$ en la ecuación (6.12):

$$\beta_{VI} = \frac{C(Y_i,\ Z_i)}{C(S_i,\ Z_i)} = \frac{C(\alpha + \beta S_i^* + e_i,\ Z_i)}{C(S_i^* + m_i,\ Z_i)}$$

Nuestra discusión de los errores en los informes recíprocos de Amalia y Amelia sobre su escolaridad presupone que $C(e_i,Z_i) = C(m_i,\ Z_i) = 0$. Esto a su vez implica que:

$$\beta_{VI} = \beta \frac{C(S_i^*,\ Z_i)}{C(S_i^*,\ Z_i)} = \beta.$$

Esta feliz conclusión procede de nuestro supuesto de que el único motivo por el que $Z_i$ está correlacionado con los ingresos es que está correlacionado a su vez con $S_i^*$. Como $S_i = S_i^* + m_i$, y como $m_i$ no tiene relación con $Z_i$, se obra el habitual milagro VI.

Po: Eso mola mazo.
*Kung Fu Panda 2*

# Abreviaturas y acrónimos

Se indica entre paréntesis la página en la que figura cada abreviatura o acrónimo.

**MC2E** mínimos cuadrados en dos etapas, un estimador de variables instrumentales que sustituye el regresor instrumentalizado por los valores estimados procedentes de la primera etapa (p. 156)

**ALS** un estudio de Joshua D. Angrist, Victor Lavy y Analia Shlosser sobre el vínculo causal entre la cantidad y la calidad de los hijos en familias israelíes (p. 150)

**BLS** Boston Latin School, la primera escuela en la jerarquía de centros selectivos de Boston (p. 189)

**C&B** College & Beyond [«Universidad y Más Allá»], un conjunto de datos (p. 69)

**FVEC** función de valor esperado condicionado, el promedio poblacional de $Y_i$ cuando $X_i$ se mantiene fijo (p. 100)

**TLC** teorema del límite central, un teorema que afirma que casi cualquier promedio muestral estará distribuido de manera normal, con una aproximación cuya precisión mejora a medida que crece el tamaño de la muestra (p. 55)

**DD** diferencias en diferencias, una herramienta econométrica que compara cambios temporales en los grupos de tratamiento y de control (p. 205)

**HIE** Health Insurance Experiment [«Experimento de Seguros Médicos»], un gran ensayo aleatorio efectuado por la RAND Corporation que asignó a las familias tratadas seguros de salud con diferentes coberturas (p. 31)

**IT** efecto de intención de tratamiento, el efecto causal promedio de una oferta de tratamiento (p. 142)

**VI** variables instrumentales, una herramienta econométrica utilizada para eliminar el sesgo de variables omitidas o el sesgo de atenuación debido a los errores de medida (p. 119)

**JTPA** Job Training Partnership Act, un programa de formación estadounidense que incluía evaluación aleatoria (p. 145)

**KIPP** Knowledge Is Power Program [«Programa Conocimiento Es Poder»], una red de centros educativos *charter* en Estados Unidos (p. 121)

**ELMT** efecto local medio del tratamiento, el efecto causal promedio del tratamiento sobre los sujetos cumplidores (p. 131)

**MVIL** estimador por máxima verosimilitud con información limitada, una alternativa a los mínimos cuadrados en dos etapas, menos afectado por sesgos (p. 169)

**LGN** ley de los grandes números, una ley estadística que establece que los promedios muestrales se aproximan a sus correspondientes promedios de población (valores esperados) a medida que crece la muestra (p. 27)

**MDVE** Minneapolis Domestic Violence Experiment [«Experimento de Violencia de Género de Minneapolis»], una evaluación aleatorizada de las estrategias policiales para combatir la violencia de género (p. 138)

**MLDA** *mínimum legal drinking age*, edad mínima legal para el consumo de alcohol en Estados Unidos (p. 172)

**AVM** accidentes con vehículos motorizados (p. 184)

**NHIS** National Health Interview Survey [«Encuesta Nacional de Salud»], un conjunto de datos (p. 17)

**OHP** Oregon Health Plan [«Plan de Salud de Oregón»], la versión oregoniana de Medicaid, un seguro médico cuyo acceso venía determinado en parte por un sorteo (p. 40)

**MCO** mínimos cuadrados ordinarios, el análogo muestral de los coeficientes de regresión de poblaciones; se usan MCO para estimar modelos de regresión (p. 75)

**SVO** sesgo de variables omitidas, la relación entre los coeficientes de regresiones efectuadas con modelos que incluyen conjuntos distintos de variables explicativas (p. 88)

**TDN** trimestre de nacimiento (p. 259)

**RD** diseño de regresión discontinua, una herramienta econométrica que se aplica cuando el tratamiento, la probabilidad del tratamiento o la intensidad media del tratamiento son una función discontinua conocida de una covariable (p. 172)

**SCR** suma cuadrática de residuos, el valor esperado (promedio poblacional) de los cuadrados de los residuos en un análisis de regresión (p. 105)

**TST** efecto del tratamiento sobre los tratados, el efecto causal promedio del tratamiento sobre la población tratada (p. 136)

**MCP** mínimos cuadrados ponderados, un estimador de regresión que asigna peso a las observaciones que se suman para calcular la SCR (p. 229)

# Notas empíricas

## Tablas

**Tabla 1.1** Características demográficas y estado de salud de parejas aseguradas y no aseguradas en la NHIS

**Fuente de los datos.** Los datos de la Encuesta Nacional de Salud (National Health Interview Survey, NHIS) de 2009 proceden de las Integrated Health Interview Series (IHIS)[1] y están disponibles en www.ihis.us/ihis/.

**Muestra.** La muestra empleada para elaborar esta tabla está formada por parejas casadas de edades comprendidas entre 26 y 59 años, con al menos uno de sus dos miembros trabajando.

**Definición de las variables.** El hecho de estar o no asegurado viene determinado por la variable UNINSURED («no asegurado») de la IHIS. El grado de salud se expresa en una escala de cinco niveles, donde 1 = mala, 2 = regular, 3 = buena, 4 = muy buena, 5 = excelente; esto procede de la variable HEALTH («salud»). El nivel cultural se construye a partir de la variable EDUC, y mide años de escolaridad terminados. A quien tiene estudios medios o ha superado las pruebas libres para obtener el título, se le han asignado 12 años de escolaridad. Las personas con algún curso universitario pero que no han

[1] Literalmente, «series integradas de encuestas de salud»; se trata de un banco público de datos que contiene información de cuatro décadas de Encuestas Nacionales de Salud (NHIS); las NHIS son la principal fuente de información sobre la salud de los estadounidenses, ya que este organismo ha recopilado datos sobre hábitos saludables, estado de salud y acceso a la atención sanitaria desde 1957. *(N. de la T.)*

llegado a graduarse tienen asignados 14 años de escolaridad. A los graduados universitarios se les asignan 16 años de escolaridad, y a los titulados superiores se les asignan 18 años. Los individuos empleados son los que «realizan algún trabajo remunerado» o los que «tienen trabajo pero no están en activo», tal como indica la variable EMPSTAT.

Los ingresos familiares se han construido asignando a cada intervalo de la variable ingresos de la IHIS (INCFAM07ON) los ingresos familiares medios para ese intervalo, basados en datos procedentes de Current Population Survey (CPS; «Encuesta de Población Actual») de 2010, suplemento de marzo (usando la variable FTOTVAL de la CPS). La muestra de la CPS empleada con esta finalidad omite observaciones como ingresos domésticos no positivos, así como las observaciones con peso negativo. En la CPS los ingresos se censuran al percentil 98; a los valores por encima del percentil 98 se les ha asignado 1,5 veces el valor de ese percentil.

**Notas adicionales a la tabla.** Todos los cálculos se han ponderado usando la variable PERWEIGHT. Los errores típicos robustos se muestran entre paréntesis.

**Tabla 1.3** Características demográficas y salud de base en el estudio RAND HIE

**Fuente de los datos.** Los datos del RAND HIE proceden de Joseph P. Newhouse, «RAND Health Insurance Experiment [in Metropolitan and Non-Metropolitan Areas of the United States], 1974-1982», ICPSR06439-v1, Consorcio Interuniversitario para la Investigación Política y Social, 1999. Este conjunto de datos se encuentra disponible en http://doi.org/10.3886/ICPSR06439.v1.

**Muestra.** La muestra usada para construir esta tabla consta de individuos adultos (de 14 o más años de edad), con datos válidos de inscripción, gastos y nivel académico.

**Definición de las variables.** Las variables demográficas del apartado A y las características de salud del apartado B se miden como referencia experimental de base. La puntuación del índice general de salud mide la percepción de los participantes acerca de su propio estado de salud en el momento de inscribirse en el estudio. Valores más elevados indican mejores estados subjetivos de salud, menor nivel de preocupación acerca de la salud y una percepción mayor de

resistencia ante la enfermedad. El índice de salud mental puntúa la salud mental de los sujetos a través de una combinación de medidas de la ansiedad, la depresión y el bienestar sicológico. Valores mayores indican mejor salud mental. La variable educativa mide el número de años de formación completados y sólo se define para individuos con edades iguales o superiores a 16 años. Los ingresos familiares se miden en dólares constantes de 1991.

**Notas adicionales de la tabla.** Los errores típicos que figuran entre paréntesis proceden del análisis por grupos a nivel familiar.

**Tabla 1.4** Gasto sanitario y resultados de salud del estudio RAND HIE

**Fuente de los datos.** Véase la nota de la tabla 1.3.

**Muestra.** Véase la nota de la tabla 1.3. La muestra del apartado A contiene observaciones múltiples para cada persona, procedentes de los distintos años de seguimiento en el marco del experimento.

**Definición de las variables.** Véanse las notas de la tabla 1.3. Las variables del apartado A se construyen a partir de las solicitudes administrativas de cada año, y las variables del apartado B se miden a la salida del experimento. La variable de consulta médica cuenta el número de veces que el sujeto recurrió a este servicio con profesionales de la salud incluidos en el seguro (excluyendo visitas cuyo carácter sea exclusivamente de odontología, sicoterapia o radiología / anestesiología / patología). La variable de ingreso hospitalario señala el número total de hospitalizaciones del sujeto, incluyendo ingresos por razones de salud mental. Las variables de gastos se miden en dólares constantes de 1991.

**Notas adicionales de la tabla.** Los errores típicos que figuran entre paréntesis proceden del análisis por grupos a nivel familiar.

**Tabla 1.5** Efectos del OHP sobre la cobertura sanitaria y el uso de servicios médicos

**Fuentes de los datos.** Los números de las columnas (1) y (2) proceden de Amy N. Finkelstein *et al.*, «The Oregon Health Insurance Experiment: Evidence from the First Year», *Quarterly Journal of Economics*, vol. 127, número 3, agosto de 2012, páginas 1057-1106. Nuestras cifras derivan de las originales de este modo:

fila (1) del apartado A, de la fila (1), columnas (1) y (2), de la tabla III;

fila (2) del apartado A, de la fila (1), columnas (1) y (2), de la tabla IV;
fila (1) del apartado B, de la fila (2), columnas (5) y (6), de la tabla V; y
fila (2) del apartado B, de la fila (1), columnas (1) y (2), de la tabla V.

Los números que constan en las columnas (3) y (4) se toman de Sarah L. Taubman *et al.,* «Medicaid Increases Emergency-Department Use: Evidence from Oregon's Health Insurance Experiment», *Science*, vol. 343, número 6168, 17 de enero de 2014, páginas 263-268. Nuestras cifras derivan de las originales de este modo:

fila (1), de la fila (1), columnas (1) y (2), de la tabla S7;
fila (3), de la fila (1), columnas (3) y (4), de la tabla S2;
fila (4), de la fila (1), columnas (7) y (8), de la tabla S2.

**Muestras.** Las columnas (1) y (2) del apartado A recurren a la muestra completa analizada en los datos de mortalidad y altas hospitalarias de Finkelstein *et al.* (2012). Las columnas (3) y (4) del apartado A proceden de los registros de visitas de doce unidades de urgencias del área de Portland desde el 10 de marzo de 2008 hasta el 30 de septiembre de 2009. El apartado B emplea los datos del estudio de seguimiento analizado en Finkelstein *et al.* (2012).

**Definición de las variables.** La información de la fila (1) en el apartado A procede de una variable binaria que marca la integración en Medicaid durante el periodo de estudio (desde la notificación del resultado del sorteo hasta el final de septiembre de 2009), información que se extrae de los datos administrativos de Medicaid. La fila (2) del apartado A corresponde a una variable binaria igual a 1 si el sujeto tuvo algún ingreso hospitalario (no debido a parto) desde la notificación hasta el final de agosto de 2009. Las variables de las filas (3) y (4) del apartado A indican si se visitaron o no los servicios de urgencias y cuántas veces se hizo. La fila (1) del apartado B mide el número de visitas (no debidas a parto) en los últimos seis meses. La variable de la fila (2) del apartado B es binaria e indica si al paciente se le recetó algún medicamento durante el estudio.

**Notas adicionales de la tabla.** Los errores típicos que figuran entre paréntesis proceden del análisis por grupos a nivel familiar.

**Tabla 1.6** Efectos del OHP sobre los indicadores de salud y sobre la salud financiera

**Fuentes de los datos.** Véanse las notas de la tabla 1.5. Los números de la fila (1) del apartado A de esta tabla proceden de la fila (2), columnas (1) y (2), de la tabla IX de Finkelstein *et al.* (2012). Las cifras de las columnas (3) y (4) proceden de Katherine Baicker *et al.*, «The Oregon Experiment-Effects of Medicaid on Clinical Outcomes», *New England Journal of Medicine*, vol. 368, número 18, 2 de mayo de 2013, páginas 1713-1722.

Las cifras de las columnas (3) y (4) se derivan de las columnas (1) y (2) del original del modo siguiente:

fila (2) del apartado A, de la fila (3) de la tabla S2;
fila (3) del apartado A, de la fila (2) de la tabla S2;
fila (4) del apartado A, de la fila (6) de la tabla S1;
fila (5) del apartado A, de la fila (1) de la tabla S1;
fila (1) del apartado B, de la fila (3) de la tabla S3; y
fila (2) del apartado B, de la fila (4) de la tabla S3.

Agradecemos a Amy Finkelstein y Allyson Barnett que nos proporcionaran los errores típicos (no publicados) de las estimaciones que figuran en Baicker *et al.* (2013).

**Muestras.** Las columnas (1) y (2) emplean la muestra del (primer) estudio de seguimiento analizado en Finkelstein *et al.* (2012). Las columnas (3) y (4) recurren a la muestra del (segundo) estudio de seguimiento analizado en Baicker *et al.* (2013).

**Definición de las variables.** La variable de la fila (1) del apartado A es binaria e indica si el sujeto valoró su propia salud como buena, muy buena o excelente (en comparación con regular o mala). Las filas (2) y (3) del apartado A contienen los índices SF-8 de salud física y mental. Mayor índice SF-8 significa mejor salud. La escala se normaliza de manera que la media y la desviación típica para la población estadounidense valgan 50 y 10, respectivamente, y el intervalo de valores vaya de 0 a 100. Véanse las páginas 14 a 16 del apéndice de Baicker *et al.* (2013) para una descripción de las medidas subjetivas y clínicas de la salud que se aportan en las filas (2) a (5). La variable de la fila (1) del apartado B es binaria y marca si el sujeto tuvo o no deudas por motivos de salud durante el estudio.

**Notas adicionales de la tabla.** Los errores típicos que figuran entre paréntesis proceden del análisis por grupos a nivel familiar.

**Tabla 2.2** Efectos de la universidad privada: emparejamientos de Barron's

**Fuentes de los datos.** Los datos usados para construir esta tabla se describen en el artículo de Stacy Berg Dale y Alan B. Krueger, «Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables», *Quarterly Journal of Economics*, vol. 117, número 4, noviembre de 2002, páginas 1491-1527.

Estos datos proceden del estudio College and Beyond (C&B), ligado a un estudio efectuado por Mathematica Policy Research Inc. entre 1995 y 1997, y de archivos proporcionados por el College Entrance Examination Board [Comité de Exámenes de Acceso a la Universidad] y el Higher Education Research Institute (HERI) [Instituto de Investigación sobre la Educación Superior] de la Universidad de California en Los Ángeles. Las categorías de selectividad de los centros son las de la publicación *Barron's Profiles of American Colleges 1978*, Barron's Educational Series, 1978.

**Muestra.** La muestra se compone de personas de la cohorte de acceso a la universidad de 1976 que constan en el estudio C&B y que tenían trabajo a tiempo completo en 1995. El análisis excluye a los estudiantes de las universidades que históricamente han acogido exclusivamente alumnado negro (Universidad Howard, Morehouse College, Spellman College y Universidad Xavier; véanse los detalles en las páginas 1500-1501 de Dale y Krueger (2002). La muestra se restringe aún más con el fin de que cada grupo de selectividad contenga tanto estudiantes que acudieron a universidades públicas como alumnos que cursaron estudios en centros privados.

**Definición de las variables.** La variable dependiente es el logaritmo de los ingresos anuales brutos en el año 1995. La pregunta formulada en el cuestionario del estudio C&B ofrece diez intervalos como respuesta; véase la nota al pie número 8, en las páginas 1501-1502 de Dale y Krueger (2002) para conocer los detalles exactos sobre la construcción de la variable de ingresos. La variable de grupo de solicitantes se forma emparejando a los estudiantes según la lista de categorías de centros donde cursaron solicitudes y donde fueron admitidos o rechazados (según el estudio C&B), donde las categorías de centros se basan en la medida del carácter selectivo de Barron's (véanse las páginas 1502-1503 de Dale y Krueger [2002]

para más detalles). La variable de nota individual en las pruebas de acceso a la universidad registra esa cifra dividida entre 100. Véase la página 1508 de Dale y Krueger (2002) para la definición de la variable de ingresos paternos (que se imputa teniendo en cuenta las profesiones y el nivel de estudios de los progenitores). Las variables mujer, negro, hispano, asiático, otra raza (o no consta), en el 10% mejor de su escuela secundaria, no consta nivel de secundaria y deportista, son binarias.

**Notas adicionales de la tabla.** Las regresiones se pesan de manera que la muestra sea representativa de la población de estudiantes de las instituciones cubiertas por C&B (véanse los detalles en la página 1501 de Dale y Krueger (2002)). Los errores típicos que figuran entre paréntesis proceden del análisis por grupos a nivel de centro de estudios.

**Tabla 2.3** Efectos de la universidad privada: controles según la nota media en las pruebas de acceso

**Fuentes de los datos.** Véanse las notas de la tabla 2.2.

**Muestra.** Véanse las notas de la tabla 2.2. La muestra empleada para construir esta tabla contiene todos los estudiantes de C&B, y no sólo aquellos asignados a grupos de selectividad de Barron's.

**Definición de las variables.** Véanse las notas de la tabla 2.2. La variable de nota individual en las pruebas de acceso a la universidad se construye de este modo: la puntuación media en los exámenes de acceso se calcula para cada universidad utilizando los datos HERI y promediando después sobre las universidades solicitadas por cada sujeto. La cifra se divide entre 100.

**Notas adicionales de la tabla.** Las regresiones se pesan de manera que la muestra sea representativa de la población de estudiantes de las instituciones cubiertas por C&B. Los errores típicos que figuran entre paréntesis proceden del análisis grupos a nivel de universidad.

**Tabla 2.4** Efectos del carácter selectivo de los centros: controles según la nota media en las pruebas de acceso

**Fuentes de los datos.** Véanse las notas de la tabla 2.2.

**Muestra.** Véanse las notas de la tabla 2.3.

**Definición de las variables.** Véanse las notas de la tabla 2.3. La variable de nota media de acceso al centro refleja la puntuación media

en los exámenes de acceso a la universidad (dividida entre 100) de los estudiantes que cursaron estudios en el centro correspondiente.

**Notas adicionales de la tabla.** Véanse las notas de la tabla 2.3.

**Tabla 2.5** Efectos de la universidad privada: sesgo de variables omitidas

**Fuentes de los datos.** Véanse las notas de la tabla 2.2.

**Muestra, definición de las variables y notas adicionales de la tabla.** Véanse las notas de la tabla 2.3.

**Tabla 3.1** Análisis de los sorteos KIPP

**Fuentes de los datos.** Los datos demográficos sobre los estudiantes de las escuelas públicas de Lynn proceden de Massachusetts Student Information Management System. Los datos demográficos y acerca del sorteo para los solicitantes KIPP procede de los archivos de la escuela KIPP de Lynn. Las calificaciones se extraen de los exámenes del Massachusetts Comprehensive Assessment System (MCAS) en matemáticas y en lengua inglesa. Véanse más detalles en Joshua D. Angrist *et al.*, «Who Benefits from KIPP?», *Journal of Policy Analysis and Management*, vol. 31, número 4, otoño de 2012, páginas 837-860.

**Muestra.** La muestra de la columna (1) contiene los estudiantes que cursaron quinto curso en centros públicos de Lynn entre el otoño de 2005 y la primavera de 2008. Las muestras de las columnas (2) a (5) se toman del conjunto de solicitantes de la escuela KIPP de Lynn para el ingreso en quinto o sexto curso en el mismo periodo. Se excluye a los solicitantes con hermanos ya matriculados en KIPP o que pasaron directamente a la lista de espera (véase la nota al pie número 14 de Angrist *et al.* (2012)). Las comparaciones relacionadas con los sorteos se restringen a los 371 solicitantes que aportaron los datos necesarios para el seguimiento.

**Definición de las variables.** Se usan variables binarias para marcar la pertenencia a los grupos hispano, negro, mujer, beca de comedor y matriculados en KIPP. Las notas en matemáticas y lengua para los estudiantes de cada curso se estandarizan con respecto a la población de referencia conformada por todos los estudiantes de ese curso en Massachusetts. Las puntuaciones de partida proceden de los exámenes de cuarto curso. Las puntuaciones de los resultados proceden de los cursos siguientes al que se realizaba cuando se formuló la

solicitud, es decir, las notas de quinto curso para quienes solicitaron el ingreso en KIPP mientras estaban en cuarto, y las notas de sexto para quienes lo hicieron mientras cursaban quinto.

**Notas adicionales de la tabla.** Los errores típicos robustos constan entre paréntesis.

**Tabla 3.3** Tratamientos asignados y aplicados en MDVE

**Fuentes de los datos.** Los números que constan en esta tabla proceden de la tabla 1 de Lawrence W. Sherman y Richard A. Berk, «The Specific Deterrent Effects of Arrest for Domestic Assault», *American Sociological Review*, vol. 49, número 2, abril de 1984, páginas 261-272.

**Tabla 3.4** Primeras fases del estudio cantidad-calidad

**Fuentes de los datos.** Los datos usados para construir esta tabla proceden de las muestras del 20% de microdatos para uso público de los censos israelíes de los años 1983 y 1995, enlazados con información no pública acerca de los progenitores y hermanos o hermanas tomada del registro civil. Véanse los detalles en Joshua D. Angrist, Victor Lavy y Analia Schlosser, «Multiple Experiments for the Causal Link between the Quantity and Quality of Children», *Journal of Labor Economics*, vol. 28, número 4, octubre de 2004, páginas 773-824.

**Muestra.** La muestra incluye judíos primogénitos de parto no múltiple y con edades entre 18 y 60 años. La muestra se restringe a individuos cuyas madres nacieron después de 1930 y que dieron a luz por primera vez con edades entre 15 y 45 años.

**Definición de las variables.** El instrumento parto múltiple (el segundo parto es de mellizos o gemelos) es una variable binaria igual a 1 para familias en las que el segundo parto produjo mellizos o gemelos. El instrumento igualdad de sexos (entre hermanos o hermanas) es una variable binaria igual a 1 si el primer y el segundo hijo son del mismo sexo.

**Notas adicionales de la tabla.** Aparte de una variable binaria que marca a los varones, se añaden variables explicativas binarias para el año del censo, procedencia étnica de los progenitores (origen asiático o africano, de la antigua Unión Soviética, de Europa o de América), y ausencia del mes de nacimiento, y otras variables explicativas para edad, edad de la madre, edad de la madre en el primer parto, y edad de la madre en el momento de la inmigración (si procede). Las pri-

meras fases de esta tabla tienen sus correspondientes segundas fases en las dos primeras filas de la tabla 3.5. Los errores típicos robustos constan entre paréntesis.

**Tabla 3.5** Estimaciones del equilibrio cantidad calidad mediante MCO y MC2E

**Fuentes de los datos.** Véanse las notas de la tabla 3.4.

**Muestra.** Véanse las notas de la tabla 3.4. Las estimaciones de las filas tercera y cuarta se limitan a los sujetos con edades entre 24 y 60 años en el momento del censo. En la variable de resultado correspondiente a la graduación universitaria faltan unos pocos valores adicionales.

**Definición de las variables.** Véanse las notas de la tabla 3.4. Las variables dependientes de las filas segunda, tercera y cuarta son binarias.

**Notas adicionales de la tabla.** Las variables explicativas aparecen listadas en las notas de la tabla 3.4.

**Tabla 4.1** Estimaciones de los efectos de la MLDA sobre la mortalidad con RD brusca

**Fuentes de los datos.** Los datos de mortalidad proceden de los archivos de detalles confidenciales de mortalidad del Centro Nacional de Estadísticas de Salud (NCHS, National Center for Health Statistics) para 1997-2004. Estos datos proceden de los certificados de defunción, y cubren todos los fallecimientos en Estados Unidos durante el periodo de estudio. Las estimaciones de población del denominador proceden de los censos de Estados Unidos de 1970-1990. Véanse los detalles en las páginas 166 a 169 de Christopher Carpenter y Carlos Dobkin, «The Effect of Alcohol Consumption on Mortality: Regression Discontinuity Evidence from the Minimum Drinking Age», *American Economic Journal-Applied Economics*, vol. 1, número 1, enero de 2009, páginas 164-182.

**Muestra.** La muestra se restringe a los fallecimientos de adultos jóvenes con edades comprendidas entre 19 y 22 años. Los datos utilizados consisten en promedios dentro de 48 celdas definidas por edad en intervalos de 30 días.

**Definición de las variables.** La causa de muerte es la que consta en los certificados de defunción según los datos NCHS. Las causas se dividen entre internas y externas, con estas últimas separadas en

categorías mutuamente excluyentes: homicidio, suicidio, accidente de tráfico, y otras causas externas. Una categoría separada de causas relacionadas con el alcohol incluye las muertes en las que se citaba el alcohol en el certificado de defunción. Los resultados son las tasas de mortalidad cada 100 000, donde el denominador procede de las estimaciones de población del censo.

**Notas adicionales de la tabla.** Se dan entre paréntesis los errores típicos robustos.

**Tabla 5.1** Empresas de venta al mayor: quiebras y ventas en 1929 y 1933

**Procedencia.** Las cifras de esta tabla proceden de la tabla 8 (página 1066) de Gary Richardson y William Troost, «Monetary Intervention Mitigated Banking Panics during the Great Depression: Quasi-Experimental Evidence from a Federal Reserve District Border, 1929-1933», *Journal of Political Economy*, vol. 117, número 6, diciembre de 2009, páginas 1031-1073.

**Fuentes de los datos.** Los datos proceden del censo de empresas de Estados Unidos de 1935, compilado por Richardson y Troost (2009).

**Tabla 5.2** Estimaciones mediante regresión DD de los efectos de la MLDA sobre las tasas de mortalidad

**Fuentes de los datos.** La información sobre MLDA por estado y año procede de «Minimum Purchase Age by State and Beverage, 1933-Present», DISCUS (Distilled Spirits Council of the US; Consejo de Licores Destilados de Estados Unidos), 1996; Alexander C. Wagenaar, «Legal Minimum Drinking Age Changes in the United States: 1970-1981», *Alcohol Health and Research World*, vol. 6, número 2, invierno 1981-1982, páginas 21-26; y William Du Mouchel, Allan F. Williams y Paul Zador, «Raising the Alcohol Purchase Age: Its Effects on Fatal Motor Vehicle Crashes in Twenty-Six States», *Journal of Legal Studies*, vol. 16, número 1, enero de 1987, páginas 249-266. Seguimos la codificacion de estas leyes tal y como se implementa en Karen E. Norberg, Laura J. Bierut y Richard A. Grucza, «Long-Term Effects of Minimum Drinking Age Laws on Past-Year Alcohol and Drug Use Disorders», *Alcoholism: Clinical and Experimental Research*, vol. 33, número 12, septiembre de 2009, páginas 2180-2190, donde se han corregido algunos errores menores de codificación.

La información sobre mortalidad procede de los datos sobre mortalidad por múltiples causas disponible en el Sistema Nacional de Estadísticas Vitales del Centro Nacional de Estadísticas de Salud, extraído de www.nber.org/data/mortality-data.html. Los datos de población proceden de las estimaciones intercensales de población de la Oficina del Censo de Estados Unidos, disponibles en la red. Véase:

http://www.census.gov/popest/data/state/asrh/pre-1980/tables/e7080sta.txt;

http://www.census.gov/popest/data/state/asrh/1980s/80s_st_age_sex.html; y

http://www.census.gov/popest/data/state/asrh/1990s/st_age_sex.html.

**Muestra.** El conjunto de datos usado para construir estas estimaciones contiene tasas de mortalidad para personas de entre 18 y 20 años de edad entre 1970 y 1983, por estado y por año.

**Definición de las variables.** La tasa de mortalidad mide el número de personas con edades entre 18 y 20 años que murieron en cada estado y cada año (por cada 100.000), según la causa de fallecimiento (todas las muertes, accidentes de tráfico, suicidio, y todas las causas internas). El regresor MLDA mide la fracción de personas entre 18 y 20 años de edad que son bebedores legales en cada estado y año. Esta fracción se calcula usando las fechas de cambio de la MLDA en cada estado, y tiene en cuenta las cláusulas de exclusión. El cálculo admite que los nacimientos se distribuyen uniformemente a lo largo de todo el año.

**Notas adicionales de la tabla.** Las regresiones de las columnas (3) y (4) incluyen pesos según la población de los estados con edades entre 18 y 20 años. Los errores típicos que constan entre paréntesis están por grupos a nivel estatal.

**Tabla 5.3** Estimaciones mediante regresión DD de los efectos de la MLDA, con control de los impuestos sobre la cerveza

**Fuentes de los datos.** Véanse las notas de la tabla 5.2. Los datos sobre impuestos proceden de Norberg *et al.*, «Long-Term Effects», *Alcoholism: Clinical and Experimental Research*, 2009.

**Muestra.** Véanse las notas de la tabla 5.2.

**Definición de las variables.** Véanse las notas de la tabla 5.2. Los

impuestos sobre la cerveza se miden en dólares constantes de 1982 por galón.

**Notas adicionales de la tabla.** Véanse las notas de la tabla 5.2.

**Tabla 6.2** Rendimientos de la formación académica para los gemelos de Twinsburg

**Fuentes de los datos.** Los datos sobre gemelos figuran detallados en Orley Ashenfelter y Cecilia Rouse, «Income, Schooling, and Ability: Evidence from a New Sample of Identical Twins», *Quarterly Journal of Economics*, vol. 113, número 1, febrero de 1998, páginas 253-284. Estos datos se encuentran disponibles en http://dataspace.princeton.edu/jspui/handle/88435/dsp01xg94hp567. Ese enlace incluye los datos utilizados en Orley Ashenfelter y Alan B. Krueger, «Estimates of the Economic Returns to Schooling from a New Sample of Twins», *American Economic Review*, vol. 84, número 5, diciembre de 1994, páginas 1157-1173.

**Muestra.** La muestra consiste en 680 gemelos entrevistados en el Festival de Gemelos de Twinsburg en 1991, 1992 y 1993. La muestra se restringe a gemelos residentes en Estados Unidos que hubieran tenido trabajo durante los dos años anteriores a la entrevista.

**Definición de las variables.** Las estimaciones de esta tabla se construyeron usando los datos sobre años de escolaridad especificados en los autoinformes y en los informes de los hermanos gemelos, definidos como el informe individual sobre los años de escolaridad logrados por el gemelo o la gemela.

**Notas adicionales de la tabla.** Se dan los errores típicos robustos entre paréntesis.

**Tabla 6.3** Rendimientos de la formación académica con el instrumento «trabajo infantil»

**Fuentes de los datos.** Los datos usados para construir esta tabla se detallan en Daron Acemoglu y Joshua D. Angrist, «How Large Are Human-Capital Externalities? Evidence from Compulsory-Schooling Laws», en Ben S. Bernanke y Kenneth Rogoff (editores), *NBER Macroeconomics Annual 2000*, vol. 15, MIT Press, 2001, páginas 9-59.

**Muestra.** La muestra consiste en varones blancos estadounidenses con edades comprendidas entre 40 y 49 años, entrevistados para los censos de Estados Unidos desde 1950 hasta 1990. La muestra se extra-

jo de muestras integradas de microdatos para uso público (IPUMS) de estos censos.

**Definición de las variables.** La variable dependiente es el logaritmo de los ingresos semanales. La variable de escolaridad admite como valor máximo 17. La variable de escolaridad del censo de 1990 se imputa parcialmente usando medias categoriales procedentes de otras fuentes de datos. Los instrumentos de legislación sobre trabajo infantil son variables binarias que indican la escolaridad exigida para permitir el acceso al mercado laboral en el estado de nacimiento del sujeto, de acuerdo con las leyes en vigor cuando el sujeto tenía 14 años de edad. Véanse los detalles en las páginas 22-28 y en el apéndice B de Acemoglu y Angrist (2001).

**Notas adicionales de la tabla.** Todas las regresiones están pesadas por medio de la variable de peso IPUMS. Los errores típicos que se dan entre paréntesis están agrupados al nivel de estado.

**Tabla 6.4** Receta VI para una estimación de los rendimientos de la formación académica con un solo instrumento «trimestre de nacimiento»

**Fuentes de los datos.** Los datos usados para construir esta tabla se detallan en Joshua D. Angrist y Alan B. Krueger, «Does Compulsory School Attendance Affect Schooling and Earnings?», *Quarterly Journal of Economics*, vol. 106, número 4, noviembre de 1991, páginas 979-1014.

**Muestra.** La muestra consiste en varones nacidos entre 1930 y 1939 que figuran en el 5% accesible al público del censo estadounidense de 1980. Se excluyeron del análisis las observaciones con valores completados por asignación, así como los encuestados que hicieron constar ingresos nulos y los que declararon no haber trabajado en 1979. Véanse las páginas 1011-1012 del apéndice 1 de Angrist y Krueger (1991).

**Definición de las variables.** El logaritmo de los ingresos semanales en 1979 se calcula dividiendo los ingresos anuales entre el número de semanas trabajadas. La variable de escolaridad corresponde al curso más alto completado.

**Notas adicionales de la tabla.** Los errores típicos robustos se dan entre paréntesis.

**Tabla 6.5** Rendimientos de la formación académica con instrumentos «trimestre de nacimiento» alternativos

**Fuentes, muestra, definición de las variables y notas adicionales de la tabla.** Véanse las notas de la tabla 6.4.

**Figuras**

**Figura 2.1** La FVEC y la recta de regresión

**Fuente de los datos.** Se trata de la figura 3.1.2 de la página 39 de Joshua D. Angrist y Jörn-Steffen Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press, 2009.

**Muestra.** Véanse las notas de la tabla 6.4.

**Definición de las variables.** La variable dependiente es el logaritmo de los ingresos semanales. La variable de escolaridad representa el curso más alto completado.

**Figura 3.1** Datos de solicitudes y matrículas de los sorteos KIPP en Lynn

**Fuente de los datos.** Véanse las notas de la tabla 3.1.

**Muestra.** El conjunto de datos KIPP analizados contiene las solicitudes iniciales para plazas en los cursos quinto y sexto en los años 2005 a 2008. Esta muestra contiene 446 solicitudes e incluye algunas sin los datos de partida requeridos para el seguimiento.

**Figura 3.2** VI en la escuela: el efecto de estudiar en KIPP sobre las notas en matemáticas

**Fuente de los datos.** Véanse las notas a la tabla 3.1.

**Muestra.** La muestra se corresponde con la de la columna (3) de la tabla 3.1.

**Figura 4.1** Cumpleaños y funerales

**Fuente de los datos.** Esta figura procede del apéndice A de Christopher Carpenter y Carlos Dobkin, «The Effect of Alcohol Consumption on Mortality: Regression Discontinuity Evidence from the Minimum Drinking Age», *American Economic Journal-Applied Economics*, vol. 1, número 1, enero de 2009, páginas 164-182.

**Notas adicionales de la figura.** La figura representa gráficamente el número de muertes en Estados Unidos entre 1997 y 2003 en función de la edad en días medida desde los cumpleaños.

**Figura 4.2** Estimación de los efectos de mortalidad de la MLDA mediante RD brusca

**Fuente de los datos y muestra.** Véanse las notas de la tabla 4.1.

**Definición de las variables.** Véanse las notas de la tabla 4.1. El eje *Y* mide la mortalidad (cada 100.000) por todas las causas. Los promedios de la figura corresponden a 48 celdas definidas por la edad en intervalos de 30 días.

**Figura 4.4** Control cuadrático en un modelo de RD

**Fuente de los datos, muestra y definición de las variables.** Véanse las notas de la tabla 4.1.

**Notas adicionales de la figura.** Véanse las notas de la figura 4.2.

**Figura 4.5** Estimación mediante RD del efecto de la MLDA sobre la mortalidad según causa de muerte

**Fuente de los datos y muestra.** Véanse las notas de la tabla 4.1.

**Definición de las variables.** Véanse las notas de la tabla 4.1. El eje *Y* mide la mortalidad (cada 100.000) según la causa de muerte. Los promedios de la figura corresponden a 48 celdas definidas por la edad en intervalos de 30 días.

**Notas adicionales de la figura.** Véanse las notas de la figura 4.2.

**Figura 4.6** Matriculación en BLS

**Fuente de los datos.** Esta figura utiliza los datos de las escuelas públicas de Boston sobre las solicitudes en las escuelas selectivas, incluyendo información acerca del Examen de Acceso de las Escuelas Independientes (ISEE, Independent School Entrance Exam), estado de matriculación entre 1999 y 2008, y notas obtenidas en los exámenes del Massachusetts Comprehensive Assessment System (MCAS) de los años académicos desde 1999/2000 hasta 2008/2009. Véanse los detalles en las páginas 142-143 y el apéndice C del suplemento al artículo de Atila Abdulkadiroglu, Joshua D. Angrist y Parag Pathak, «The Elite Illusion: Achievement Effects at Boston and New York Exam Schools», *Econometrica*, vol. 81, número 1, enero de 2014, páginas 137-196. El suplemento se encuentra accesible en http://www.econometricsociety.org/ecta/supmat/10266_data_description.pdf.

**Muestra.** La muestra incluye los estudiantes matriculados en escuelas públicas de Boston que solicitaron acceso en la Boston Latin

School (BLS) para plazas de séptimo curso desde 1999 hasta 2008. La muestra se restringe a los estudiantes para los que BLS era la primera opción, o bien se convirtió en la primera opción tras eliminar las escuelas en las que el estudiante no logró el acceso.

**Definición de las variables.** La variable móvil, etiquetada como «puntuación en el examen de acceso» en la figura, es una media pesada de la puntuación total ISEE (Examen de Acceso de las Escuelas Independientes, Independent School Entrance Exam) del solicitante y su GPA (Grade Point Average, la media de sus notas finales en cada curso). La tasa de matriculación en escuelas selectivas se calcula usando datos del curso escolar posterior al de la solicitud.

**Notas adicionales de la figura.** Los valores de la variable móvil de la figura se normalizaron sustrayendo la nota más baja a la que se ofreció plaza, de manera que el umbral de todos los años resulta igual a cero. Las líneas suavizadas de las figuras son valores ajustados por medio de modelos de regresión estimados en el entorno de cada punto. Estos modelos resuelven una regresión de la variable dependiente sobre la variable móvil para las observaciones contenidas dentro de un ancho de banda no paramétrico. Véanse más detalles en Abdulkadiroglu *et al.* (2014).

**Figura 4.7** Matriculación en cualquier escuela selectiva de Boston

**Fuente de los datos, muestra y notas adicionales de la figura.** Véanse las notas de la figura 4.6.

**Definición de las variables.** Véanse las notas de la figura 4.6. La matriculación en cualquier centro selectivo indica si un solicitante se matriculó en el centro Boston Latin School, en Boston Latin Academy o en John D. O'Bryant High School of Mathematics and Science.

**Figura 4.8** Calidad de los pares en el entorno del umbral BLS

**Fuente de los datos, muestra y notas adicionales de la figura.** Véanse las notas de la figura 4.6.

**Definición de las variables.** Véanse las notas de la figura 4.6. Para cada solicitante de acceso en los centros selectivos, la calidad de los pares es el promedio de las notas de matemáticas del Massachusetts Comprehensive Assessment System (MCAS) que obtuvieron en cuarto curso sus compañeros de clase en séptimo curso, con independencia del centro en el que el solicitante cursara estudios entonces.

**Figura 4.9** Notas en matemáticas en el entorno del umbral BLS

**Fuente de los datos, muestra y notas adicionales de la figura.** Véanse las notas de la figura 4.6.

**Definición de las variables.** Véanse las notas de la figura 4.6. La variable en el eje *Y* es el promedio de las notas de matemáticas del Massachusetts Comprehensive Assessment System (MCAS) en séptimo y octavo cursos.

**Figura 4.10** RD visual de Thistlethwaite y Campbell

**Fuente de los datos.** Se trata de la figura 3 de Donald L. Thistlethwaite y Donald T. Campbell, «Regression-Discontinuity Analysis: An Alternative to the ex post facto Experiment», *Journal of Educational Psychology*, vol. 51, número 6, diciembre de 1960, páginas 309-317.

**Muestra.** La muestra contiene 5.126 casi ganadores y 2.848 casi perdedores del Certificado al Mérito en la competición de 1957 de las Becas Nacionales al Mérito. La variable móvil es la nota en el examen de calificación para la beca del comité examinador de acceso a la universidad, ahora conocido como PSAT. Las dos medidas de resultados proceden de un estudio efectuado sobre todos los estudiantes de la muestra, unos seis meses después de que se anunciara la resolución de concesión de los premios.

**Definición de las variables.** Las dos variables de resultados son variables binarias que marcan para cada estudiante si planeaba cursar tres o más años de estudios de posgrado (representado como la línea I-I'), o si planeaba convertirse en profesor universitario o investigador científico (línea J-J').

**Figura 5.1** Quiebras bancarias en los distritos sexto y octavo de la Reserva Federal

**Fuente de los datos.** Los datos diarios sobre el número de bancos activos en Misisipi fueron compilados por Gary Richardson y William Troost, y se describen en las páginas 1034-1038 de Gary Richardson y William Troost, «Monetary Intervention Mitigated Banking Panics during the Great Depression: Quasi-Experimental Evidence from a Federal Reserve District Border, 1929-1933», *Journal of Political Economy*, vol. 117, número 6, diciembre de 2009, páginas 1031-1073.

**Muestra.** Los datos de operaciones bancarias registran todos los bancos con autorización para operar a nivel nacional o estatal en Misisipi, sumados dentro de cada distrito de la Reserva Federal, y en funcionamiento el 1 de julio de 1930 y el 1 de julio de 1931.

**Definición de las variables.** El eje *Y* muestra el número de bancos abiertos y operativos el 1 de julio de cada año, en cada distrito.

**Figura 5.2** Tendencias de las quiebras bancarias en los distritos sexto y octavo de la Reserva Federal

**Fuente de los datos.** Véanse las notas de la figura 5.1.

**Muestra.** Los datos de operaciones bancarias registran todos los bancos con autorización para operar a nivel nacional o estatal en Misisipi, sumados dentro de cada distrito de la Reserva Federal, y en funcionamiento entre julio de 1929 y julio de 1934.

**Definición de las variables.** Véanse las notas de la figura 5.1.

**Figura 5.3** Tendencias de las quiebras bancarias en los distritos sexto y octavo de la Reserva Federal, y estimación DD contrafactual para el distrito sexto

**Fuente de los datos.** Véanse las notas de la figura 5.1.

**Muestra.** Véanse las notas de la figura 5.2.

**Figura 5.7** La receta DD de John Snow

**Fuente.** Se trata de la reproducción de la tabla XII (en la página 90) del libro de John Snow, *On the Mode of Communication of Cholera*, segunda edición, John Churchill, 1855.

**Figura 6.1** Primera etapa del trimestre de nacimiento

**Fuente de los datos, muestra y definición de las variables.** Véanse las notas de la tabla 6.4.

**Figura 6.2** Forma reducida del trimestre de nacimiento

**Fuente de los datos, muestra y definición de las variables.** Véanse las notas de la tabla 6.4.

**Figura 6.3** Notas en el examen de última convocatoria y titulaciones en Tejas

**Fuente de los datos.** Esta figura se ha construido a partir de un

conjunto de datos que cruza archivos administrativos de centros de enseñanza secundaria con registros administrativos de escolarización posterior a secundaria, e ingresos procedentes de seguros de desempleo en Tejas. Estos datos se detallan en las páginas 288-289 de Damon Clark y Paco Martorell, «The Signaling Value of a High School Diploma», *Journal of Political Economy*, vol. 122, número 2, abril de 2014, páginas 282-318.

**Muestra.** La muestra consiste en cinco cohortes de veteranos de la escuela secundaria que realizaron el examen de última convocatoria en las primaveras de 1993 a 1997. Los datos de ingresos están disponibles hasta 2004, es decir, por un periodo de entre 7 y 11 años tras la fecha del examen.

**Definición de las variables.** La variable móvil en el eje *X* mide la nota en el examen de última convocatoria, centrada en la puntuación de corte para el aprobado. Como el examen abarca muchas asignaturas y los aspirantes deben superarlas todas para lograr el título, las calificaciones se normalizan respecto de los umbrales de corte, y la variable móvil viene dada por el mínimo de estas notas normalizadas. El eje *Y* representa la probabilidad de obtener el título para cada valor de la nota.

**Figura 6.4.** Efectos de la nota en el examen de última convocatoria sobre los ingresos

**Fuente de los datos y muestra.** Véanse las notas de la figura 6.3.

**Definición de las variables.** La variable móvil en el eje *X* es la misma de la figura 6.3. El eje *Y* mide los ingresos anuales promedio, incluyendo ceros para quienes no trabajaban, para cada valor de la nota.

# Figuras


1.1 Una distribución normal gaussiana ..... 56
1.2 La distribución *t* para la media de una muestra de tamaño 10 ..... 57
1.3 La distribución *t* para la media de una muestra de tamaño 40 ..... 58
1.4 La distribución *t* para la media de una muestra de tamaño 100 ..... 58
2.1 La FVEC y la recta de regresión ..... 101
2.2 La varianza en *X* es buena ..... 115
3.1 Datos de solicitudes y matrículas de los sorteos KIPP en Lynn ..... 125
3.2 VI en la escuela: el efecto de estudiar en KIPP sobre las notas en matemáticas ..... 130
4.1 Cumpleaños y funerales ..... 173
4.2 Estimación de los efectos de mortalidad de la MLDA mediante RD brusca ..... 174
4.3 RD en acción, tres enfoques ..... 179
4.4 Control cuadrático en un modelo de RD ..... 183
4.5 Estimación mediante RD del efecto de la MLDA sobre la mortalidad según causa de muerte ..... 186
4.6 Matriculación en BLS ..... 190
4.7 Matriculación en cualquier escuela selectiva de Boston ..... 191
4.8 Calidad de los pares en el entorno del umbral BLS ..... 193
4.9 Notas en matemáticas en el entorno del umbral BLS ..... 198
4.10 RD visual de Thistlethwaite y Campbell ..... 203
5.1 Quiebras bancarias en los distritos sexto y octavo de la Reserva Federal ..... 212
5.2 Tendencias de las quiebras bancarias en los distritos sexto y octavo de la Reserva Federal ..... 213
5.3 Tendencias de las quiebras bancarias en los distritos sexto y octavo de la Reserva Federal, y estimación DD contrafactual para el distrito sexto ..... 213

5.4 Un efecto MLDA en estados con tendencias paralelas....226
5.5 Un efecto MLDA espurio en estados con tendencias no paralelas....226
5.6 Un efecto MLDA real, visible aunque las tendencias no sean paralelas....227
5.7 La receta DD de John Snow....234
6.1 Primera etapa del trimestre de nacimiento....260
6.2 Forma reducida del trimestre de nacimiento....260
6.3 Notas en el examen de última convocatoria y titulaciones en Tejas....267
6.4 Efectos de la nota en el examen de última convocatoria sobre los ingresos....267

# Tablas


1.1 Características demográficas y estado de salud de parejas aseguradas y no aseguradas en la NHIS ..... 20
1.2 Resultados y tratamientos para Khuzdar y María ..... 21
1.3 Características demográficas y salud de base en el estudio RAND HIE ..... 35
1.4 Gasto sanitario y resultados de salud del estudio RAND HIE ..... 38
1.5 Efectos del OHP sobre la cobertura sanitaria y el uso de servicios médicos ..... 42
1.6 Efectos del OHP sobre los indicadores de salud y sobre la salud financiera ..... 44
2.1 Matriz universitaria de emparejamiento ..... 70
2.2 Efectos de la universidad privada: emparejamientos de Barron's ..... 81
2.3 Efectos de la universidad privada: controles según la nota media en las pruebas de acceso ..... 84
2.4 Efectos del carácter selectivo de los centros: controles según la nota media en las pruebas de acceso ..... 85
2.5 Efectos de la universidad privada: sesgo de variables omitidas ..... 95
3.1 Análisis de los sorteos KIPP ..... 126
3.2 Los cuatro tipos de muchachos ..... 134
3.3 Tratamientos asignados y aplicados en MDVE ..... 139
3.4 Primeras fases del estudio cantidad-calidad ..... 159
3.5 Estimaciones del equilibrio cantidad calidad mediante MCO y MC2E ..... 161
4.1 Estimaciones de los efectos de la MLDA sobre la mortalidad con RD brusca ..... 185
5.1 Empresas de venta al mayor: quiebras y ventas en 1929 y 1933 ..... 218
5.2 Estimaciones mediante regresión DD de los efectos de la MLDA sobre las tasas de mortalidad ..... 224

5.3 Estimaciones mediante regresión DD de los efectos de la MLDA, con control de los impuestos sobre la cerveza ..... 229
6.1 Un mal control genera sesgos de selección ..... 245
6.2 Rendimientos de la formación académica para los gemelos de Twinsburg ..... 250
6.3 Rendimientos de la educación con el instrumento «trabajo infantil» ..... 256
6.4 Receta VI para una estimación de los rendimientos de la formación académica con un solo instrumento «trimestre de nacimiento» ..... 261
6.5 Rendimientos de la formación académica con instrumentos «trimestre de nacimiento» alternativos ..... 262

# Agradecimientos

Georg Graetz, Kyle Greenberg, Christian Pérez, Miikka Rokkanen, Daisy Sun, Chris Walters y Alicia Xiong contribuyeron con investigaciones experimentadas. Noam Angrist, A. J. Bostian, Stephanie Cheng, Don Cox, Dan Fetter, Yi Jie Gwee, Samuel Huang, Ayrat Maksyutov, Thomas Pischke y Melvyn Weeks nos aportaron lecturas concienzudas y comentarios escritos. Nos sentimos especialmente agradecidos con Gabriel Kreindler, a quien debemos la meticulosa recopilación de las notas empíricas, y con Mayara Silva, por su atenta revisión e inestimable organización del manuscrito final. Los casos prácticos se prepararon gracias a la colaboración de los indulgentes maestros Kitt Carpenter, Damon Clark, Stacy Dale, Carlos Dobkin, Amy Finkelstein, Karen Norberg, Gary Richardson y Analia Schlosser, a quienes agradecemos tanto su ayuda como sus datos. También merecen una mención especial nuestro editor, Seth Ditchik de Princeton University Press, por alentar y dirigir este proyecto; nuestros diestros y disciplinados editores de producción, Princeton Editorial Associates y Terri O'Prey en la editorial; así como Garrett Scafani y Yeti Technologies por la formidable originalidad de las ilustraciones.

En esta empresa, como en todo lo demás, nuestros seres más queridos nos alumbran el camino.

# Índice analítico

Los números de página que remiten a entradas dentro de figuras van seguidos de una efe (f); los que remiten a entradas que constan en notas, por una ene (n); y los que remiten a entradas que aparecen dentro de tablas, por una te (t).


Abdulkadiroglu, Atila, 192*n*, 196*n*
ACA (Affordable Care Act, Ley de Atención Médica Asequible), 15, 39
accidentes con vehículos motorizados (AVM), muertes relacionadas con el alcohol, 184-185, 185*t*, 186*f*, 223, 224*t*
accidentes. *Véase* accidentes de tráfico
Acemoglu, Daron, 254, 254*n*, 255, 259
acta nacional sobre la edad mínima legal para el consumo de alcohol, 219
Affordable Care Act (ACA, Ley de Atención Médica Asequible), 15, 39
agresión marital: Experimento de Violencia de Género de Minneapolis (Minneapolis Domestic Violence Experiment), 138, 139*t*; caso Simpson, 137
Alabama, edad mínima legal para el consumo de alcohol en, 220-221
alcohol:
  impuestos sobre, 228-229, 229*t*
  precios de la bebida, 228
  y accidentes de tráfico, 183-185, 185*t*, 186*f*, 223, 224*t*
  y mortalidad, 172-189, 219-231
  *Véase además* edad mínima legal para el consumo de alcohol
Altonji, Joseph, 97*n*
análisis de camino, 164
ancho de banda, 187, 189
Angrist, Joshua D., 102*n*, 123*n*, 135*n*, 141*n*, 145*n*, 150*n*, 152*n*, 157*n*, 192*n*, 254*n*, 255, 259, 259*n*, 264*n*. *Véase además* maestro Joshway
apéndice B de *The Tariff on Animal and Vegetable Oils* (Wright), 163-164
Arkansas, edad mínima legal para el consumo de alcohol en, 216
Armada británica, 46
Aron-Dine, Aviva, 32*n*, 33*n*
Ashenfelter, Orley, 247, 247*n*, 251
asignación aleatoria:
  de los grupos de control y de tratamiento, xiii, 27-31

diferencias con el muestreo aleatorio, 49
elimina el sesgo de selección, 30
*Véase además* sorteos
atención médica:
gastos en, 15, 37, 38*t*, 44*t*, 43
elasticidad de los precios ante la demanda de, 32
Atlanta, banco de la Reserva Federal en, 208-219
atrición, 33n, 127n
autocorrelación, 233-236
autocorrelación, 233-235
AVM. *Véase* accidentes con vehículos motorizados
ayuda financiera, xiii, 65-66

Bagehot, Walter, 208, 208*n*
Baicker, Katherine, 41*n*
Baker, Regina M., 263*n*
banco central. *Véase* Reserva Federal
bancos de la Reserva Federal:
Atlanta, 208-219
San Luis, 208-219
bancos:
quiebras en la Gran Depresión, 205-206, 209-216, 212, 213*f*
liquidez de, 208-209
Banerjee, Abhijit, 33*n*
bebida. *Véase* alcohol; edad mínima legal para el consumo de alcohol
Becker, Gary S., 147, 147*n*
Berk, Richard A., 138*n*
Bloom, Howard S., 145*n*
BLS. *Véase* Boston Latin School
Bongaarts, John, 148*n*
Boston Latin Academy, 189, 193
Boston Latin School (BLS), 189, 190*f*, 191*f*, 192-194, 193*f*, 196-197, 198*f*
Boston, escuelas selectivas en, 189-190, 192-194, 191*f*, 199
Bound, John, 263*n*
brecha educativa, 121, 194-195
Brook, Robert H., 32*n*
Buckles, Kasey, 263, 263*n*

C&B (conjunto de datos College and Beyond), 69, 76-77, 87
Caldwell and Company, 206, 209, 214
Caldwell, Rogers, 206-207
Campbell, Donald T., 201-202, 201*n*, 203*n*,
capital humano. *Véase* formación académica
características demográficas, 34, 35*t*
Carnoy, Martin, 122*n*
Carpenter, Christopher, 173*n*, 189*n*, 219*n*
categorías selectivas de Barron's, 77, 77*n*
*ceteris paribus*:
e inferencia causal, 10-14
en regresiones, 64-67, 86-93. *Véase además* sesgo de selección
Cheng, Shaohua, 299
China, política del hijo único, 147
Clark, Damon, 266, 266*n*
Clark, Kerry, 164-165, 165*n*
Coale, Ansley, 148*n*
cociente intelectual (CI) y rendimientos de la formación académica, 242
cólera, 232-233, 234*f*
College and Beyond (C&B, Universidad y Más Allá), conjunto de datos, 69, 76-77, 87
comprobación del equilibrio, 31, 34, 124
conducción bajo los efectos del alcohol. *Véase* accidentes con vehículos motorizados
control malo, 243-246, 245*t*
Cook, Thomas D., 203, 203n
coste de oportunidad de la formación académica, 265
covarianza, 104-105. *Véase además* varianza
crecimiento de población, 146-147.

*Véase además* tamaño familiar
crecimiento económico, 146-147
crisis financiera:
riesgo moral en, 208
semejanzas entre diversas, 207. *Véase además* bancos
cumpleaños:
tasas de mortalidad en el vigésimo primer, 172-176, 173*f*, 174*f*, 178, 181-189
trimestre de nacimiento, 258-265, 260*f*
cumplidores, 133, 133-137
cumplimiento en experimentos aleatorios, 143-145
curva de demanda, 163-164
curva de oferta y demanda, 163-164

Dale, Stacy Berg, 68, 68*n*, 69, 86, 86*n*
Daniel, Libro de, 46-47
Darwin, Charles, 98-99
datos tabulados, 222, 233
DD. *Véase* diferencias en diferencias
*Desmadre a la americana*, 172
desviación típica muestral combinada, 61
desviación típica muestral, 54, 61
desviación típica poblacional, 51-52
desviación típica:
muestral, 51, 53
muestral combinada, 61
poblacional, 51
diferencias de ingresos por género, 67
diferencias en diferencias (DD):
ejemplo de las quiebras bancarias, 208-219
suposición de tendencias comunes en, 212-214, 224-225
contrafactual, 211-214, 213*f*
ejemplo de la edad mínima legal para el consumo de alcohol, 220-229, 224*t*
ejemplo de la política monetaria, 208-219
modelos de regresión para, 214-216, 220-229
estimaciones de los rendimientos de la formación académica, 254-257, 256*t*
errores típicos en, 233-236
efectos estatales en, 221-223
con tendencias estatales específicas, 224-229, 226*f*, 227*f*
efectos temporales en, 220, 220*n*
diseño de regresión discontinua brusca, 175-188, 174*f*, 194
diseño de regresión discontinua difusa, 192-200, 265-268
diseño de regresión discontinua no paramétrica, 186-188
diseño de regresión discontinua paramétrica, 185
distribución normal estándar, 71-72, 56*f*, 56-57
distribución:
normal estándar, 54, 56*f*, 57
de variables, 52
Dobkin, Carlos, 173*n*, 189*n*, 219*n*
doctrina «Real Bills», 209
Duflo, Esther, 33*n*

edad mínima legal para el consumo de alcohol (MLDA, *minimum legal drinking age*):
análisis mediante regresión discontinua, 175-189
análisis mediante diferencias en diferencias, 220-230, 224*t*
en leyes federales y estatales, 219-220, 228
y mortalidad, 172-176, 182-188, 220-231, 224*t*, 229*t*
educación pública. *Véase* escuelas *charter*; formación académica

efecto causal, 16-20. *Véase además* efecto causal promedio; efecto de intención de tratamiento; efecto local medio del tratamiento; efecto del tratamiento sobre los tratados
efecto causal promedio, 22-24
efecto de la intención de tratamiento, 142-143
efecto local medio del tratamiento, 131-137
en el diseño RD, 181-183
efecto del tratamiento sobre los tratados, 136, 143-145
efecto de intención de tratamiento (IT), 142-143
efecto de los pares en la formación académica, 83, 86, 190, 194-198, 193*f*, 198*f*
efecto del tratamiento sobre los tratados (TST), 136-137, 144-145
efecto local medio del tratamiento (ELMT), 131-137
definición, 131
estimación, 132
validez externa, 136-137
efectos anuales. *Véase* efectos temporales
efectos badana, 265-268, 237f
efectos del tratamiento. *Véase* efecto causal promedio; efecto de intención de tratamiento; efecto local medio del tratamiento; efecto del tratamiento sobre los tratados
efectos estatales, 221, 222, 253-257
efectos temporales, 220, 220*n*, 222-223, 222*n*
Ehrlich, Paul, 146, 147
Einav, Liran, 32*n*, 33*n*
elasticidad de los precios con la demanda para la atención médica, 32
identificación, 163
Elder, Todd, 97*n*
Ellement, John R., 68*n*
ELMT. *Véase* efecto local medio del tratamiento
emparejamiento *(matching)*, 67-68, 69-70, 70*t*
empleo. *Véanse* ingresos; profesión
empresas de venta al mayor, 217, 218*t*
ensayo aleatorio:
ventajas, 11-14, 26, 28-31
análisis, 26-31
sobre los efectos de los seguros médicos, 26-27, 31-39, 40-45
historia, 45-48
con cumplimento imperfecto, 138-145
sobre respuesta policial a la violencia de género, 138-145
muestras para, 28-29
epidemiología, 232-233
equilibrio cantidad-calidad. *Véase* tamaño familiar
equilibrio racial en colegios públicos, 199
equilibrio, comprobación del, 31, 34-37, 124
error de medida, 249-252, 264, 269-274
error típico estimado, 54, 61. *Véase además* error típico
error típico robusto por grupos, 235
error típico robusto, 116
error típico: por grupos, 235
para la comparación de medias, 31, 60
definición, 54
para diferencias en diferencias, 233-236
estimado, 54, 61
con variables instrumentales, 167
en regresiones, 80, 114-116
robusto, 117
y el tamaño de la muestra, 116
y la significanción estadística, 36, 57-58

en mínimos cuadrados en dos etapas, 154
error. *Véase* error de medida; error típico
escorbuto, 46
escuela secundaria:
exámenes finales, 266-268, 267*f*
escuelas selectivas, 189-192, 195-200. *Véase además* formación académica
escuelas *charter*:
debate sobre ellas, 120-122
definición, 120
enfoques educativos, 120-121, 137
KIPP, 120-137
notas de examen en, 124-128, 126*t*, 130*f*
escuelas selectivas:
calidad de los pares en, 189, 192-195, 193*f*, 195-199, 198*f*
composición racial, 199-200
diferencias con escuelas públicas no selectivas, 199-200
exámenes de ingreso, 189-192, 197
modelo de autorrevelado, 83-84, 84*t*, 86, 96
umbrales de admisión, 189-193, 190*f*, 191*f*
escuelas. *Véase* escuelas *charter*; formación académica; escuelas secundarias
estados de Estados Unidos:
leyes de trabajo infantil, 254, 255, 256*t*
leyes de escolarización obligatoria, 252-257, 258
expansión de Medicaid en, 39-45. *Véase además* edad mínima legal para el consumo de alcohol
estandarización de calificaciones:
en escuelas *charter*, 124-128, 126*t*, 129-130, 130*f*
definición, 124-125
en exámenes finales de secundaria, 266-268, 267*f*
en las notas de acceso a la universidad, 67, 66, 66*n*, 69, 82, 95-97, 202
en los exámenes de admisión para escuelas selectivas, 189-194, 197
estatura, 98
estilometría, 164
estimador estadístico:
definición, 50
no sesgado, 50
estimador no sesgado, 50
estimador por máxima verosimilitud con información limitada (MVIL), 168-169
estudiantes de minorías:
brecha educativa, 121, 194
en escuelas *charter*, 120, 121, 122
en escuelas selectivas, 199-200
y fin de la segregación racial educativa, 199-200
estudiantes. *Véase* formación académica; estudiantes de minorías
eugenesia, 48, 98-99
Evans, William, 152, 152*n*
experiencia potencial, 239, 240, 246*n*
experimento de campo:
análisis con variables instrumentales, 140-145
Minneapolis Domestic Violence Experiment (Experimento de Violencia de Género de Minneapolis), 138-145, 139*t*
experimento RAND HIE (Health Insurance Experiment), 31-39, 33*n*, 35*t*, 38*t*, 43-45, 48-49. *Véase además* ensayo aleatorio
experimento natural, 172, 176-177, 232

*F*, estimador estadístico, 168-169, 262*t*, 263

facultades. *Véase* universidad; formación académica
falta de datos. *Véase* atrición
fertilidad. *Véase* planificación familiar; tamaño familiar
fiabilidad, 271
Finkelstein, Amy, 32*n*, 33*n*, 41, 41*n*
Fisher, Ronald A., 47-48, 47*n*, 164
forma reducida, variables instrumentales, 131, 132, 136, 141-142
  en diseños de regresión discontinua difusa, 196-200
  en el ejemplo de los rendimientos de la formación académica, 252, 255, 256*t*, 259, 260*f*, 266-268
  en mínimos cuadrados en dos etapas, 154, 155-158, 167, 168-170
formación académica superior. *Véase* formación académica
formación académica, rendimientos de la:
  control de aptitud, 241, 246
  control de experiencia laboral, 239, 246*n*
  control de profesión, 243-246, 245*t*
  efectos badana, 265-268, 267*f*
  efectos de la titulación, 265-268, 267*f*
  estimación mediante variables instrumentales, 252-265, 256*t*, 261*t*, 262*t*
  estimaciones mediante modelos, 246-252, 250*t*
  estimaciones por diferencias en diferencias, 255-257, 256*t*
  estimaciones por regresión, 238-246, 247-251, 255-256, 262
  función de valor esperado condicionado, 100, 101*f*, 102-103
  para gemelos, 246-252, 250*t*
  y costes de oportunidad, 265
  y errores de medida, 249-252, 264
  y leyes de escolarización obligatoria, 252-257, 256*t*
  y sesgo de aptitud, 241, 246-248
  y trimestre de nacimiento, 257-265, 260*f*, 261*t*, 262*t*
formación académica:
  brecha educativa, 121, 194
  calidad de las universidades 63-86
  calidad de los pares en, 193-199, 193f, 198f
  deuda estudiantil, xii-xiii; formación universitaria, 63-64, 65.
  escuelas charter, 120-122
  escuelas selectivas, 189-195, 194-200
  fin de la segregación racial, 198-200
  políticas de asignación de centro, 194
  programa de Becas al Mérito Nacional, 201-202, 203f
  Véase además formación académica, rendimientos de la
  y tamaño familiar, 148-162
fórmula del sesgo de variables omitidas (SVO), 88-93, 95-97, 110-113, 164, 241
Friedman, Milton, 207, 207*n*, 218, 218*n*
Friedman, Rose D., 218, 218*n*
Frost, Robert, 20*n*
  «El camino no tomado», 17
función de valor esperado condicionado (FVEC), 100-103, 101*f*
función núcleo *(kernel)* uniforme, 187*n*
FVEC (función de valor esperado condicionado), 100-103, 101*f*

Gage, Nathaniel L., 203*n*
Gallagher, Hugh, 87
Galton, Francis, 48, 98-100, 99*n*, 164
gemelos o mellizos:
  como instrumento para el tamaño familiar, 149-151, 159-162

rendimientos de la formación académica para, 246-252, 250*t*
Gladwin, Bertie, 237, 238, 265
Goldman, Ronald, 138, 145
Gran Depresión:
quiebras bancarias en, 205-206, 210-216, 212*f*, 213*f*
y política monetaria, 206-219
Griliches, Zvi, 242, 242n
Gruber, Jonathan, 16*n*
grupo de comparación. *Véase* grupo de control
grupo de control:
comprobación del equilibrio, 31, 34-37, 124
definición, 17-18. *Véase además* asignación aleatoria
grupo de riesgo, 128
grupo de tratamiento:
comprobación del equilibrio, 31, 34-37
definición, 18. *Véase además* asignación aleatoria

Haldane, J.B.S., 47
Harmenberg, Johan, 241, 242
Health Insurance Experiment (HIE, Experimento de Seguros Médicos), 31-39, 33*n*, 35*t*, 38*t*, 43-45, 48-49
heredabilidad, 98-99
hermanos, igualdad de sexos entre, 152-154, 157-162, 159*t*, 161*t*. *Véase además* gemelos o mellizos
heterocedástico, 117
HIE. *Véase* Health Insurance Experiment
hipótesis de efectos constantes, 24
hipótesis de trabajo, 55
hipótesis nula, 55
hipótesis nula, 55
homocedástico, 116
Hungerman, Daniel M., 263, 263*n*

Imbens, Guido W., 135*n*, 157*n*, 188*n*,
impuesto sobre la cerveza, 228-229, 229*t*
impuestos sobre el alcohol, 228-229, 229*t*
independencia estadística, 53
Independent Schools Entrance Exam (ISEE), 192-193, 197
India:
nivel de vida en, 147, 147*n*
planificación familiar en, 147-148
indicador estadístico (muestral): definición, 50; error típico, 54
inferencia causal, 45, 63, 66
inferencia estadística, 48
ingresos:
diferencias de género, 67
potenciales, 242, 253
y años de experiencia laboral, 239, 240*n*. *Véase además* formación académica, rendimientos de la
Iniciativa Amatista, 172, 180
Instituto de la Cerveza, 228
instrumento, definición, 128
intervalo de confianza, 57, 61
ISEE (Independent Schools Entrance Exam), 192-193, 197
Israel:
estudio ALS, 150-153, 158-159
demografía de, 149-150
IT. *Véase* efecto de intención de tratamiento

Jacobsen, Rebecca, 122*n*
Jaeger, David A., 263*n*
Jagger, Mick, 241, 242
Jalil, Andrew, 236*n*
Jan, Tracy, 68*n*
Jastrow, Joseph, 47, 47*n*
Job Training Partnership Act (JTPA), 145n

Kalyanaraman, Karthik, 188*n*
Knowledge Is Power Program (KIPP, Programa Conocimiento es Poder) en escuelas *charter*, 120-122, 123-130, 132-133, 136-137
Krueger, Alan B., 68, 68*n*, 69, 86, 86*n*, 247, 247*n*, 251, 259, 259*n*, 264*n*
*Kung Fu*, 9-10, 39, 48, 63, 119, 171, 189, 205,
*Kung Fu Panda*, 13, 257
*Kung Fu Panda 2*, 219, 274,

Lam, David, 146*n*
Lavy, Victor, 150, 150*n*
Lewis, H. Gregg, 147*n*
ley de los grandes números (LGN), 28-32
leyes de escolarización obligatoria, 252-257
leyes inglesas de pobreza (English Poor Laws), 99
LGN (ley de los grandes números), 28-32
licor. *Véase* alcohol; edad mínima legal para el consumo de alcohol
Lind, James, 46-47, 47*n*
logros académicos:
  de las madres, 263-264
  de las mujeres, 148, 151n, 263-264
  efectos badana, 266-268, 267*f*
  equilibrio de sexos entre hermanos, 152
  graduación en secundaria, 266-268, 267*f*
  según el trimestre de nacimiento, 259, 260*f*, 261, 261*t*
  y estado de salud, 19-21
Lombard College, 165
Londres, epidemias de cólera en, 232-233, 234*f*
Lutz, Wolfgang, 146*n*
Lynn, Massachusetts, 122, 130. *Véase además* Knowledge Is Power Program en escuelas *charter*

madres:
  de gemelos o mellizos, 150-151, 151*n*, 154-155. *Véase además* tamaño familiar
  logros académicos de las, 151*n*, 263-264
  por fecundación *in vitro*, 151*n*
maestro Joshway, 45, 98, 123, 145, 146, 150, 152, 162, 243, 254, 257
maestro Stevefu, 97, 98, 200, 231, 265, 268
Malthus, Thomas, 146
Martorell, Paco, 266, 266*n*
Marx, Groucho, 64
Mathews, Jay, 121*n*
Matthau, Walter, 26
MC2E. *Véase* mínimos cuadrados en dos etapas
MCO (mínimos cuadrados ordinarios), 75, 75*n*, 169-170
MCP. *Véase* mínimos cuadrados ponderados
MDVE (Minneapolis Domestic Violence Experiment, Experimento de Violencia de Género de Minneapolis), 138-145, 139*t*
media poblacional. *Véase* valor esperado matemático
media: 23-24, 49-50
  diferencia en, 23-26, 58-60
  poblacional, 50. *Véanse además* regresión a la media, media muestral
medias muestrales:
  carácter no sesgado, 50
  diferencias de, 23-26, 58-60
  errores típicos de, 54-55
  estimación de las medias poblacionales a partir de, 50

estimador estadístico *t* para, 55-59, 57*f*, 58*f*
muestreo de la distribución de, 55-59, 57*f*, 58*f*
varianza de muestreo de, 52-54
Medicaid, 16, 39-45, 44*t*
Medicare, 16
Mincer, Jacob, 238-242, 238*n*, 248
mínimos cuadrados en dos etapas (MC2E): 254-162, 166-170
ejemplo del tamaño familiar, 155-161, 161*t*
el problema de la pluralidad de instrumentos débiles, 169-170
errores típicos, 159, 167
forma reducida, 154, 155-156, 166, 168-170
primera etapa, 154, 158, 159-160, 159*t*, 166-167
segunda etapa, 156, 157-158, 160, 161, 166-167
variables de control en, 156
mínimos cuadrados ordinarios (MCO), 75, 75*n*, 168-170
mínimos cuadrados ponderados (MCP), 187*n*, 229-231
Minneapolis Domestic Violence Experiment (MDVE, Experimento de Violencia de Género de Minneapolis), 138-145, 139*t*
Mishel, Lawrence, 122*n*
Misisipi:
empresas de venta al mayor en, 217, 218*t*
frontera entre distritos de la Reserva Federal en, 209
quiebras de bancos en, 206, 209-215, 212*f*, 213*f*
MLDA. *Véase* edad mínima legal para el consumo de alcohol
modelos de ecuaciones simultáneas, 163-164
monotonía, 135
Morris, Carl, 33*n*
mortalidad:
en el vigésimo primer cumpleaños, 172-174, 173*f*, 174*f*, 182-188
por cólera, 232-233, 234*f*
según causa de muerte, 183-187, 185*t*, 186*f*
y edad mínima legal para el consumo de alcohol, 172-177, 182-188, 220-231, 224*t*, 229*t*
mortalidad. *Véase* tasa de mortalidad
muestreo aleatorio, 49, 53
mujeres: ingresos, 66-67
logros académicos, 148, 151*n*, 263-264. *Véase además* madres

Nabucodonosor, rey, 46
natalidad. *Véase* planificación familiar; tamaño familiar
National Health Interview Survey (NHIS, Encuesta Nacional de Salud), 17-21, 20*t*, 25
National Merit Scholarship program (programa de Becas al Mérito Nacional), 201-203, 203*f*
Newhouse, Joseph P., 37*n*
NHIS (National Health Interview Survey, Encuesta Nacional de Salud), 17-21, 20*t*, 25
niños:
edades de escolarización, 258-259, 264
igualdad de sexos entre hermanos, 152-154, 157-162, 159*t*, 161*t*. *Véase además* escuelas *charter* formación académica tamaño familiar; estudiantes de minorías; gemelos o mellizos; tipos de hijos
nivel de vida y tamaño familiar, 148-149. *Véase además* crecimiento económico; pobreza

notas de acceso a la universidad, 64, 67, 66*n*, 69, 82, 94-97, 202
Nueva York, escuelas selectivas en la ciudad de, 189-190
nunca tomadores, 133, 134

OHP. *Véase* Oregon Health Plan (Plan de Salud de Oregón)
Oregon Health Plan (OHP, Plan de Salud de Oregón), sorteo, 40-45, 42*t*, 44*t*
Orr, Larry L., 145*n*

parámetro poblacional, 24, 50-51
parámetro, 24, 50-51
Pathak, Parag, 192*n*
Pearson, Karl, 99
Peirce, Charles S., 47, 47*n*
peso asignado por población, 230-231
peso con funciones núcleo (*kernel functions*), 187*n*
Pingle, Robert, 138
Pischke, Jörn-Steffen, 102*n*, 145*n*. *Véase además* maestro Stevefu
planificación familiar, 147-148, 148*n*
Platón, 191
pobreza: infantil, 121, 122-123
  leyes inglesas de pobreza (English Poor Laws), 99
  segregación residencial por, 194. *Véase además* nivel de vida; Medicaid
policía. *Véase* Minneapolis Domestic Violence Experiment (Experimento de Violencia de Género de Minneapolis)
política monetaria:
  y actividad económica, 216-218, 218*t*
  en la Gran Depresión, 206-219
  doctrina «Real Bills», 209
primera etapa, variables instrumentales, 125*f*, 128-129, 131, 132, 134, 136, 140-141
  en diseños de regresión discontinua difusa, 197-199
  en el ejemplo de rendimientos de la formación académica, 252, 254-255, 256*t*, 259, 260*f*, 266-268
  en mínimos cuadrados en dos etapas, 154, 155-157, 159*t*, 166-167
probabilidad, 50
problema de identificación, 163-164
profesión y rendimientos de la formación académica, 243-246, 245*t*
promedio muestral. *Véase* medias muestrales
promedio poblacional. *Véase* valor esperado matemático
promedio ponderado por regresión, 73, 75-76
promedio. *Véase* valor esperado matemático; media; media muestral
puntos logarítmicos, 114

RAND Health Insurance Experiment. *Véase* Health Insurance Experiment
RD. *Véase* regresión discontinua
regresión a la media, 99
regresión con diferencias en diferencias (DD):
  ejemplo de la edad mínima legal para el consumo de alcohol, 220-223
  ejemplo de las quiebras bancarias, 214-216
  errores típicos en, 233-236
regresión discontinua (RD): 175-178
  ancho de banda, 187, 189
  brusca, 174-189, 174*f*, 195
  centrado de la variable móvil, 176

comparación con la regresión, 177
con control cuadrático de la variable móvil, 180-182, 183*f*
con términos de interacción, 180, 181
difusa, 191-200, 265-268
no paramétrica, 187-189
paramétrica, 186
variable móvil, 176, 177, 178
visual, 183
y comportamientos no lineales, 178-182, 179*f*
regresión lineal local, 190*n*
regresión múltiple, 108-112. *Véase además* sesgo de variables omitidas
regresión simple, 94-94, 104-105
regresión: 73-76, 102-105
análisis de sensibilidad, 93-97
anatomía, 108-110
bivariada, 104-105, 105*n*, 108
coeficientes, 74
con logaritmos, 212-213
errores de medida en, 269-274
errores típicos, 114-217
larga, 88-89, 90, 91-92, 94-97, 98
lineal local, 192*n*
multivariada, 105*n*, 108-112
pesada 229-231. *Véase además* variables de control
residuos, 75, 105-106
sesgo de variables omitidas en, 88-97, 95*t*, 110-113
simple, 88, 90-91, 92, 98
uso por Galton, 98-99
valores estimados, 106
variable dependiente, 73-74
variables binarias en, 74, 107-108, 109-110
y emparejamiento (*matching*), 73-74, 75-76
y función de valor esperado condicionado, 100-104, 101*f*
Reinhart, Carmen, 207, 207*n*
rendimientos de la formación académica. *Véase* formación académica, rendimientos de la
Reserva Federal:
distrito octavo, 209-219, 212*f*, 213*f*
distrito sexto, 209-219, 212*f*, 213*f*
política monetaria de, 208-219
residuos:
autocorrelación, 233
cuadráticos, 75
definición, 75
en mínimos cuadrados en dos etapas, 168
propiedades, 106
restricción de exclusión, 123, 128, 143, 153-154
resultados potenciales, 19-20, 242
resultados:
definición, 18
observados, 20-22
potenciales, 20-22, 242
pretratamiento, 36-37
variable de, 73, 88, 131. *Véase además* variable dependiente
retadores, 135-135
Richardson, Gary, 209, 209*n*
riesgo moral, 208
Rimer, Sara, 138*n*
Rogoff, Kenneth, 207, 207*n*
Rosenzweig, Mark R., 149, 149*n*
Rothstein, Richard, 122*n*
Rouse, Cecilia, 247, 247*n*, 251
Rubin, Donald B., 135*n*

salarios. *Véase* ingresos; formación académica, rendimientos de la
Salter, James, 17
San Luis, bando de la Reserva Federal en, 209-219
Sandburg, Carl, 165
Sanderson, Warren, 146*n*
Scherbov, Sergei, 146*n*

Schlosser, Analia, 150, 150*n*
Schwartz, Anna, 207, 207*n*
SCR (suma cuadrática de residuos), 105, 230
seguros médicos:
  Affordable Care Act (Ley de Atención Médica Asequible), 15
  beneficios financieros de, 44*t*, 43-45
  efectos de los ensayos aleatorios, 27-28, 31-39, 40-45
  Medicaid, 16, 39-45
  Medicare, 16
  nacionales, 21
  relación con la salud, 15-24, 20*t*, 37-39, 38*t*, 42-45, 44*t*
  sujetos no asegurados, 16-17, 19-21, 20*t*, 21-22, 39
  y el uso de la atención médica, 37, 38*t*, 41-42, 42*t*, 43-45
seguros. *Véase* seguros médicos
sesgo de aptitud, 240-243
sesgo de atenuación, 251-252, 271-272
sesgo de muestra finita en MC2E, 169-170
sesgo de selección: 12-14
  debido a mal control, 243-246, 245*t*
  definición, 12
  en el Experimento de Violencia de Género de Minneapolis, 138-139, 143-144
  en la asistencia a escuelas *charter*, 125-127, 130-131
  en la elección de centro universitario, 64-65, 71, 86, 87-88
  en la estimación de los rendimientos de la formación académica, 239-241
  en mínimos cuadrados en dos etapas, 168-170. *Véase además* sesgo de variables omitidas
  y diferencias de medias, 24-25
sesgo de variables omitidas (SVO), 88-97, 95*t*, 111-112, 155, 177, 200, 241-241. *Véase además* sesgo de selección
Sherman, Lawrence W., 138*n*
siempre tomadores, 133-134, 144-145, 153
significancia estadística, 36, 56-58, 59-60, 61-62
significación. *Véase* significación estadística
Simpson, Nicole Brown, 138
Simpson, O. J., 137, 138*n*
Snow, John, 232-233, 233*n*, 234*f*
sorteos de escuelas *charter*, 122-123
  análisis mediante variables instrumentales de, 122-130, 126*t*
  escuelas *charter*, 122-130
  sorteos, Oregon Health Plan (Plan de Salud de Oregón), 40-45, 42*t*, 44*t*
Stanley, Julian C, 202, 203*n*
Stock, James H., 164, 164*n*, 165*n*
suicidio, 184, 185*t*, 223
suma cuadrática de residuos (SCR), 105, 230
suposición de independencia, 128-129
SVO. *Véase* sesgo de variables omitidas

*t*, indicador estadístico:
  definición, 55
  distribución de muestreo de, 55-59, 57*f*, 58*f*
  para la comparación de medias, 61
  para la media muestral, 55
tabaco, 47-48
Taber, Christopher, 97*n*
tabla estado-año, 223, 230, 231
tamaño familiar:
  estudio ALS, 150-153, 158-160
  y capital humano de los hijos, 148-162
  y elección de universidad, 87-93

y nivel de vida, 147-148
reducción del, 147-148, 148*n*
e igualdad de sexos entre hermanos, 152-154, 147-162, 159*t*, 161*t*
tamaño muestral:
y distribuciones de muestreo, 55-56, 57*f*, 58*f*
y estimadores no sesgados, 50
y la asignación aleatoria, 27-28
y la ley de los grandes números, 27-31
y la varianza de muestreo, 53-54
y los errores típicos, 53-54, 114
Taubman, Sarah, 41*n*
TDN. *Véase* trimestre de nacimiento
Teach for America, 121, 121*n*
Tejas, exámenes finales de secundaria en, 266-268, 267*f*
tendencias comunes,
suposición, 211-214
relajación, 223-228
tendencias paralelas. *Véase* suposición de tendencias comunes
Tennessee, edad mínima legal para el consumo de alcohol en, 220
teorema del límite central (TLC), 55
término de error. *Véase* residuos
término de interacción, 180-182, 215-216
Thistlethwaite, Donald L., 201-202, 202*n*
Thomas, Duncan P., 47*n*
tipos de hijos, 133, 134
TLC (teorema del límite central), 55-57
Tomes, Nigel, 147*n*
trabajo infantil, leyes, 254, 255, 256*t*
transición demográfica, 148n. *Véase además* tamaño familiar
tratamiento, definición, 18
Trebbi, Francesco, 164, 164*n*
Tribunal Supremo de Estados Unidos, 199
trimestre de nacimiento (TDN), rendimientos de la formación académica empleando, 258-264, 260*f*, 261*t*, 262*t*
Troost, William, 209, 209*n*
Twinsburg, Ohio, 246-252

universidad:
ayuda financiera para, 12, 65-66
privada frente a pública, 63-97. *Véase además* formación académica

validez externa, 137
valor esperado condicionado, 29-30, 100
valor esperado matemático, 28, 49-51
valor esperado:
condicionado, 29-30, 100
matemático, 28, 50-51
valor estimado, 75, 106-107
variabilidad, medida de la, 51-54
variable binaria, 23, 29-30, 74, 78, 79, 107-108, 109-110
variable de control:
buena, 246
definición, 73-74
en mínimos cuadrados en dos etapas, 156-158. *Véase además* sesgo de variables omitidas
mala, 243-246
variable de tratamiento,
definición, 18
en regresión discontinua, 175, 192-195
en regresiones, 73, 74
para análisis con variables instrumentales, 131
para análisis de diferencias en diferencias, 220
variable dependiente:
definición, 73-74
logarítmica, 78, 112-114. *Véase además* resultado

valores estimados, 106-107
variable móvil, 176, 177, 178
variable:
dependiente, 73-74, 78, 105-106, 112-113
distribución de, 52
binaria 23, 29-30, 74, 78, 79, 107-108, 108-109
móvil, 176, 177, 178
de tratamiento, 18, 73, 74, 131, 192-195, 220. *Véase además* variable de control; variables instrumentales; resultados
variables explicativas, 108-112, 251, 262, 271-272
variables instrumentales (VI):
como reacción en cadena, 129, 131-132, 135
efecto local medio del tratamiento, 131-137
ejemplo de los sorteos de las escuelas *charter*, 122-130, 133-137
ejemplo del experimento de Violencia de Género de Minnesota (Minnesota Domestic Violence Experiment), 138-145
ejemplo del tamaño familiar, 150-153
elimina el sesgo de selección, 143
en experimentos de campo, 140-145
errores típicos, 167. *Véase además* formación académica, rendimientos de la
estimador, 119, 128, 167
inferencia causal basada en, 123, 127-128, 131-133, 135, 154
invención del método, 163-165
y error de medida, 273-274
variación de muestreo. *Véase* varianza de muestreo
varianza de muestreo, 36, 53-54, 60-61
varianza muestral, 51, 52*n*, 60-61
varianza poblacional, 51, 52-53
varianza: definición, 51-52
de muestreo, 48, 52-54, 60, 61-62, 114-117. *Véase además* covarianza
de población, 51, 52
descriptiva, 52
diferencias en, 60-65, 117
muestral, 51-52, 52*n*, 60-61
residual, 114-117
VI. *Véase* variables instrumentales
violencia de género. *Véase* Minneapolis Domestic Violence Experiment; agresión marital
Virtue, G. O., 163*n*

*Waiting for Superman*, 120
Wheeler, Adam, 68*n*
Wilcox, Moses y Aaron, 246
Wolpin, Kenneth I., 149, 149*n*
Wright, Philip G., 163-165, 163*n*, 165*n*
Wright, Sewall, 47, 163

Yule, George Udny, 98-99, 99*n*

Zappa, Frank. *Véase* Platón

**«Pocas ramas de la estadística han avanzado tanto en las últimas décadas como la inferencia causal. El estilo atractivo y revelador de Angrist y Pischke descubre a los lectores cinco métodos poderosos de esta materia. Si aspira a buscar relaciones de causa-efecto, o a entender los resultados obtenidos por otros, querrá leer este libro cuanto antes.»**

GARY KING
Harvard University

**«Centrándose en cinco herramientas econométricas, esta obra expone conceptos clave de esta rama de la economía y será de utilidad en cualquier disciplina que emplee técnicas estadísticas para realizar inferencias causales.»**

MELVYN WEEKS
University of Cambridge

**«Los autores recurren a ejemplos fascinantes para transmitir a sus lectores la intuición y el conocimiento estadístico que necesitan para dominar la econometría. ¡Una lectura obligada para cualquiera que trabaje con datos!»**

MELISSA S. KEARNEY
The Brookings Institution y University of Maryland

La econometría aplicada es la ciencia de los datos en su estado original, y engloba los métodos estadísticos que se usan en economía para desentrañar causas y efectos de las actividades humanas. Con un lenguaje accesible y algunas dosis de humor con sabor a kung-fu, esta obra expone las herramientas esenciales del análisis econométrico y desvela por qué la econometría es una disciplina tan apasionante y útil.

¿Mejoran la salud los seguros médicos? ¿Son preferibles las universidades de élite que otro tipo de centros de enseñanza superior? Cuando la banca privada se tambalea, y los inversores toman el dinero y huyen, ¿deben acudir en su ayuda los bancos centrales? Angrist y Pischke nos muestran que, con los métodos adecuados, la econometría es capaz de ofrecernos respuestas a preguntas tan dispares como estas.

ISBN 978-84-941595-0-3

Antoni Bosch editor

www.antonibosch.com

9 788494 159503